THE MATHEMATICAL-STATISTICAL METHODOLOGY OF THE CONTEMPORARY SOVIET FAMILY BUDGET SURVEY

Stephen Shenfield

A thesis submitted for the degree of

Doctor of Philosophy

Centre for Russian and East European Studies
Faculty of Commerce and Social Science
The University of Birmingham
1984

Vol 1 of 2

ACKNOWLEDGEMENTS

First of all I am indebted to my supervisor, Dr. Phil Hanson, both for his extremely detailed and helpful comments on successive drafts of this thesis and for encouraging me to return to the thesis when my attention strayed for too long to other matters. I have also been encouraged by the interest which the late Dr. John Grayson, and Drs. Alastair McAuley and Ann Helgeson, have taken in my research.

I would also like to thank Mrs.Jenny Brine of the Baykov Library for all her help with library matters, and Mrs. Olga Griffin for her speedy typing of the final draft.

I am grateful to the Science and Engineering Research Council and to the University of Birmingham for financial support.

Synopsis

The study describes and assesses the mathematical-statistical methodology of the contemporary Soviet Family Budget Survey, both in regard to the sample design and in regard to the processing and analysis of the survey data. A wide range of methodological deficiencies are identified, accounting for the widely recognised unreliability of the data produced.

The problems of using the survey data in various fields of policy-making, planning and research are explored. It is shown that Soviet data-users where possible avoid relying on data from this survey.

The historical and social factors influencing the methodology of the Family Budget Survey are discussed. The most important causes of the deficient methodology are found to be the neglect of mathematical statistics and sampling theory in Soviet socio-economic statistics, originating in the Stalin period, and the bureaucratic inertia of the Central Statistical Administration.

TABLE OF CONTENTS

Table of Contents cont'd..

: : : : : : : :

PART A

INTRODUCTION AND BACKGROUND

CHAPTER A1

SCOPE, PURPOSE AND CONTENT OF THESIS

This thesis is a study of the methodology[1] of the Soviet Family Budget Survey (FBS) conducted by the Central Statistical Administration (TsSU)[2] and of the role it plays in Soviet planning and policy-making. Although we shall consider the influence of earlier practice on the contemporary survey, our concern is with the budget survey in the form it has taken since its reorganisation in 1951-2.

The FBS is the most extensive continuous[3] sample survey conducted in the USSR. On it "the State spends many millions of roubles annually" (Karapetyan 1980 p.127). It is, moreover, an extremely labour-intensive operation, occupying about 18 per cent of the staff of TsSU offices (Chapter A2). It is also the only source of some important types of economic data on the way of life of the Soviet population (Chapter D5). However, it has not previously been examined in depth by Western researchers, though it is discussed briefly by Kaser (1955), Hanson (1968 pp.78-9), Goldman (1972 p.321) and McAuley (1979 pp.51-53). The lack of attention the FBS has received may be explained by the fact that most researchers have been interested in assessing published Soviet data rather than in the Soviet statistical system as a phenomenon worthy of study in its own right. Very few data indeed from the FBS are published.

A report on the FBS has been published by the Foreign Demographic Analysis Division of the Bureau of the Census of the US Department of Commerce (DePauw 1965). This is a useful source of detailed information on the forms used by survey interviewers, on the calculations carried out by them, and on the TsSU Work Plan which specifies the reports based on

FBS data required from oblast' statistical offices. In our view, however, DePauw is insufficiently critical of the survey methodology. It will be seen that we do not accept his argument that "the needs of the planners, administrators and researchers for periodic detailed information on the level of living and the daily economic activities of the Soviet people are probably well met by the programme" (p.2).

In this study we make no attempt to discuss each and every item of information collected in the FBS. Our purpose is rather to assess the basic methodology of the survey by the standards of sampling theory and mathematical statistics, and to consider the consequences of the methodology for data quality and usability[4]. Finally, we aim to place TsSU practice in this field in its social and historical context.

We draw comparisons between the Soviet FBS and the budget surveys of Western and of other East European countries where it is especially useful to do so, but there is no intention to describe systematically any surveys except the Soviet one.

The thesis consists of five parts. Part A provides background necessary for an understanding of the rest of the thesis. Part B deals with sample design, both the existing design of the survey and its representativeness and possible alternative designs. In Part C we consider various important aspects of the processing and analysis of survey data, and in Part D we discuss the problems of using survey data in a number of important fields of research, planning and policy-making. In Part E we conclude with a discussion of the implications of our findings for the nature of the Soviet statistical system and, more generally, for the nature of Soviet society.

Chapter A2 provides some necessary information about the organisational structure of the Central Statistical Administration. In Chapter A3 we place the subject of the thesis in its historical context by reviewing

aspects of the development of the discipline of mathematical statistics, and of the branch of that discipline constituted by modern sampling theory, in the West, in pre-revolutionary Russia and in the Soviet Union. We shall see that the question of the application of the methods of mathematical statistics to socio-economic statistics has been surrounded by controversy in the USSR since Stalinist times. The position that sampling occupies in contemporary Soviet statistics, and the historically conditioned deficiencies in contemporary Soviet sampling practice, are examined in Chapter A4. The main stress is laid on those deficiencies which also affect the design of the FBS.

We begin our treatment of the sample design of the FBS in Chapter B1 with a critical account of its general principles and of the biases entailed by them. In Chapter B2 we consider how the total size of the sample, and its composition by "social groups" (workers, employees, collective farmers, pensioners), have changed over time. The incomplete territorial coverage of the sample is analysed in Chapter B3, and its incomplete branch and occupational coverage in Chapter B4. The representativeness of survey data is affected not only by the factors dealt with in Chapters B1-B4 but also by the consequences of the practice of inducing families to participate in the survey for prolonged periods. This factor, and others associated with participation in the survey, are discussed in Chapter B5, completing the basic description of the sample design.

Certain standard checks are carried out by TsSU on the representativeness of sections of the FBS sample. These checks and their deficiencies are considered in Chapter B6. We then, to the extent that it is possible to do so, make a general assessment of the representativeness of the sample, and especially of its income distribution, in Chapter B7. We conclude that the sample is subject to a great many different biases, often severe and cumulative in effect, and that the survey data are therefore

highly unrepresentative of the population as a whole.

TsSU is aware that FBS data are unrepresentative, and makes limited efforts to reduce the biases in the sample. These efforts are assessed in Chapter B8. Data users often attempt to compensate for biases in the data by applying to them corrective coefficients estimated with the help of other, more reliable, sources of statistical information, an expedient discussed in Chapter B9. The technique is shown to have a modest potential for extracting, under certain conditions, some useful results from biased data.

The two main proposals for the radical reconstruction of the FBS sample advanced by Soviet writers are discussed in Chapters B10 and B11. Many writers urge that selection on the economic branch principle be abandoned and the survey be consistently reorganised on the territorial principle, as is the practice in all Western and some East European countries. It is also often suggested that observation of families over prolonged periods of time should be fully or partially replaced by "the method of momentary observations". This method, developed by Soviet researchers, involves the collection of different categories of budget data over periods of different length, mostly very short (one day, one week etc.).

In some respects the sample design of the contemporary budget survey can be accounted for in terms of the legacy of the early "monographic" budget surveys, dating from the nineteenth century. We consider this point in Chapter B12.

While our treatment of the issues associated with the sample design is as complete as the available literature permits, we do not aim at similar completeness in our discussion of the processing and analysis of FBS data. In particular, we do not attempt to describe all known work on FBS data carried out in various institutes outside TsSU. In Chapters

C1-C4 we review the processing and analysis of the collected data in TsSU offices and computational installations, examining in Chapter C5 a specific aspect of the data-processing.

Chapter C1 describes the preliminary processing and checking of collected data carried out by the staff of oblast' statistical offices responsible for budget survey work. This reveals further deficiencies in the quality of the data eventually produced from the FBS.

Chapters C2-C4 are devoted to the main bulk of the data-processing work regularly performed by TsSU at oblast', Republic and All-Union levels, and to the continuing process of first mechanisation and then computer-isation of data-processing. Chapter C2 provides a background account of the technical characteristics of the data-processing equipment at the disposal of TsSU at different times, from primitive keying machines to the automated data banks presently being introduced. We discuss how technical development affects the practicability of different modes of data analysis, increasing the scope for flexibility and the use of the methods of mathematical statistics.

In Chapters C3 and C4 we review the data-processing programmes actually in operation or planned for the future, noting the limited extent to which this potential has so far been exploited in practice. In Chapter C3 we consider the production by TsSU of statistical tabulations from FBS data, concentrating on the selection of criteria for grouping family budgets together. The constraint which the traditional procedure of "decentralised summarisation" imposes on flexible tabulation is noted. Other parts of the TsSU data-processing programme, and in particular those involving the use of mathematical-statistical methods, are considered in Chapter C4.

Chapter C5 is devoted to the problem of specifying product nomenclatures for the tabulation of consumption data.

In Chapters D1-D4 we consider various fields of research, planning and

policy-making for which the FBS is a possible source of statistical data. We examine the extent to which FBS data are in fact used and the way in which they are used, on their own or in combination with data from other sources. The possible reasons for the limited use made of FBS data - poor data quality, secrecy, theoretical opposition to methods based on them, and so on - are discussed.

In Chapter D1 we attempt a broad survey of the use made of FBS data and of how this use has changed over time. Special attention is paid to the role played by FBS data in the making of welfare policy. In Chapter D2 we deal with the use made of FBS data in the construction of a number of statistical and planning balances, such as the balances of money incomes and expenditures of the population and the balances of agricultural production. The relationship between such balances and tasks like the planning of monetary circulation is explained. Chapter D3 is concerned with the choice of data sources, mainly between FBS data and the data of retail trade statistics, for the study and forecasting of consumer demand. In Chapter D4 we consider the FBS as an unused source of data on the consumption of alcoholic drink.

On the basis of our survey of the use made of FBS data and of the specific case-studies examined, we make in Chapter D5 a general assessment of the problems of using FBS data for practical purposes.

In Part E we analyse the underlying causes of the methodological inertia characteristic of the conduct of the contemporary Family Budget Survey within a broader social context. The problems of innovation in the methodology of the FBS are compared with those of technological innovation in Soviet industry and of reform in Soviet environmental policy, as well as with the general problem of overcoming the ideological deformation of Soviet science.

Notes to Chapter A1

1 Thus we are not directly concerned with the subject-matter about which the FBS collects data, nor with analysing the small quantity of published FBS data available. For a preliminary summary of the thesis, see Shenfield (1983b).

2 Budget surveys on a smaller scale have been conducted by the Soviet Trade Union Council (VTsSPS) and by various research institutes. Some of these are mentioned in the thesis, but no attempt is made to discuss them systematically.

3 The survey of family composition, incomes and living conditions conducted by TsSU every three years uses a sample considerably larger than that of the FBS, but has a much more restricted programme of observation.

4 Thus there are quite a few important issues of substantive (rather than mathematical) methodology, such as the assessment of the value of socially subsidised services consumed by the population, which we do not discuss.

CHAPTER A2

ORGANISATION OF THE CENTRAL STATISTICAL ADMINISTRATION

1 Introduction

The Family Budget Survey is conducted by the Budget Statistics Department (Otdel byudzhetnoi statistiki; OBS[1]) of the Central Statistical Administration (Tsentralnoe statisticheskoe upravlenie; TsSU), the main body responsible for statistical work in the USSR[2]. The purpose of this chapter is to provide necessary background information on the organisation and staffing of TsSU as a whole (Section 2) and of the Department of Budget Statistics in particular (Section 3)[3].

2 Organisation and staffing of TsSU

TsSU is subordinate to the USSR Council of Ministers, of which its director, L.M.Volodarskii (previously, V.N.Starovskii), is a member. TsSU is managed by a group of its senior staff called its Collegium (Kollegiya). The Collegium is advised by a Scientific-Methodological Council (Nauchno-metodologicheskii sovyet) of academic statisticians, which does not appear to exercise any great influence.

TsSU has under its jurisdiction various educational, research and publishing bodies (Table A2.1). These include two scientific research institutes:

(a) the Scientific Research Institute of TsSU (Nauchno-issledovatelskii institut TsSU; NII TsSU), concerned with a wide range of statistical research. Its director is at present a specialist in input-output analysis, M.R.Eidelman; the previous director was the socio-demographic statistician A.Ya.Boyarskii. NII TsSU has branches (filialy)

TABLE A2.1

ORGANISATIONS SUBORDINATE TO THE CENTRAL STATISTICAL ADMINISTRATION (TsSU)

Educational (1)

About 12 statistical tekhnikumy (colleges of secondary specialised education) for training lower level statistical staff

About 300 accountancy tekhnikumy for training industrial, agricultural and administrative accountants (Yezhov 1967)

Research and advisory

Scientific-Methodological Council

TsSU Scientific Research Institute (NII TsSU)

TsSU All-Union State Design Institute for the Mechanisation of Accounting and Computational Work (VGPTI TsSU)

Publishing (2)

Publishing House "Statistika" (formerly "Gosstatizdat")

The monthly journal Vestnik statistiki (Statistical Courier)

Notes: (1) Higher level statistical staff are trained in higher educational institutions - in the statistics faculties of economics institutes, in the statistics kafedry of the economics faculties of universities, and in the Moscow Economic-Statistical Institute (MESI). These are all subordinated to the Ministry of Higher and Special Secondary Education, though TsSU has attempted to gain control over MESI.

(2) TsSU is supplied with statistical forms by the organisation Soyuzuchetizdat.

in the Union Republics as well as a central All-Union organisation (NII TsSU USSR); the Latvian branch has carried out research into budget survey methodology (Chapter B11);

(b) the All-Union State Design Institute for the Mechanisation of Accounting and Computational Work (Vsesoyuznyi gosudarstvennyi proyektno-tekhnicheskii institut TsSU; VGPTI TsSU), which designs computer systems for TsSU's network of computer centres.

TsSU is subdivided both vertically, into offices at different geographical levels, and horizontally, into departments fulfilling different functions.

The vertical subdivision of TsSU

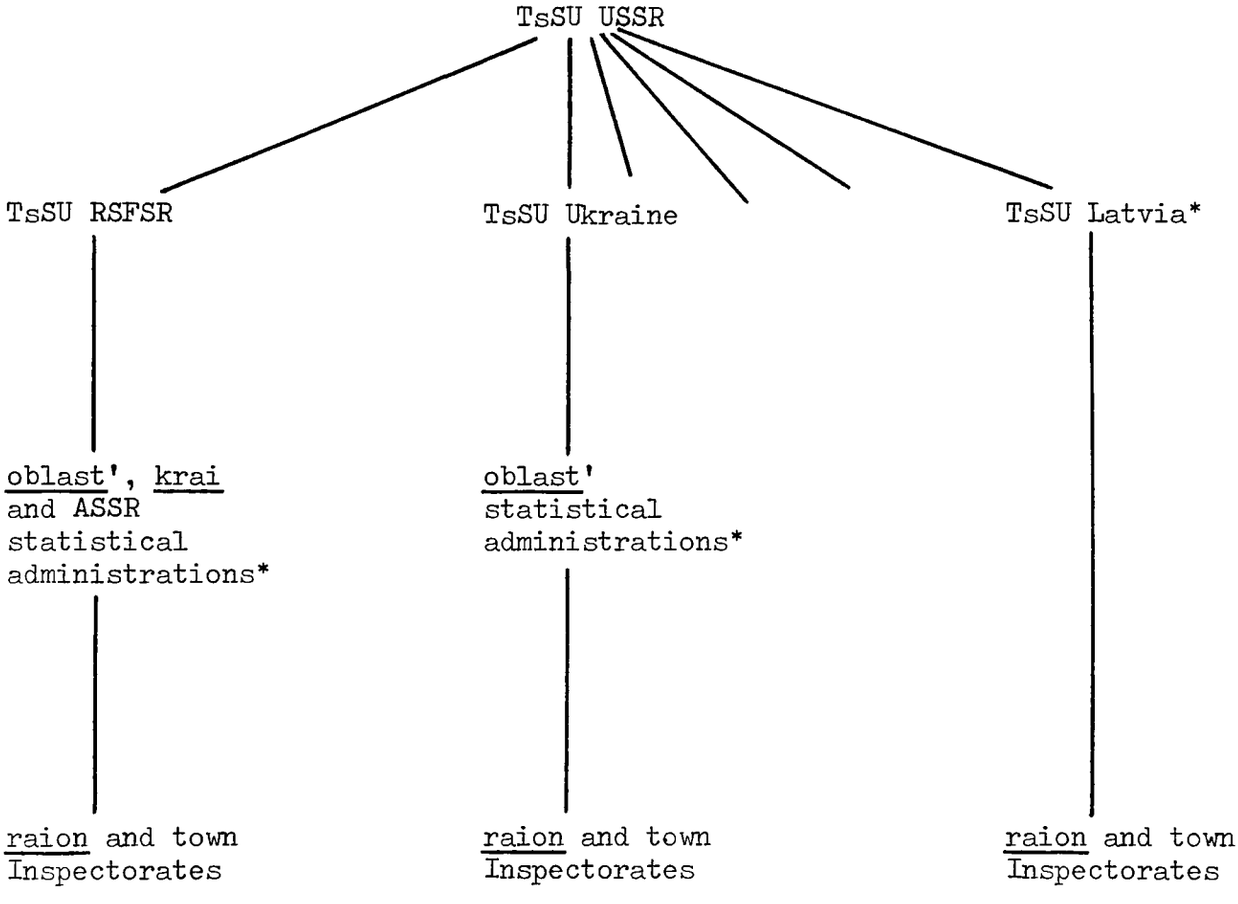
Vertical subdivisions of TsSU correspond to Soviet territorial-administrative units (Table A2.2). Beneath the central All-Union organisation, TsSU USSR, there is an organisation for each Union Republic: TsSU RSFSR, TsSU Ukrainian SSR, TsSU Latvian SSR and so on. The statistical administrations (statisticheskie upravleniya or statupravleniya) of oblasti, krai, autonomous republics (ASSR's) and large cities (Moscow, Leningrad etc.) are subordinate to the TsSU's of the "large" Republics which are divided into such units (RSFSR, the Ukraine, Belorussia etc.). The lowest level of the hierarchy is constituted by the Town and raion Inspectorates of State Statistics (gorodskie /raionnye inspektury gosudarstvennoi statistiki) in urban and rural areas respectively, subordinate to the statistical administrations of oblasti, krai and ASSR's and to the TsSU's of "small" Republics without oblast' subdivision (Latvia, Moldavia etc.).

It will be convenient to refer to all units immediately above raion level - oblasti, krai, ASSR's, large cities and small Republics - as "oblast'-level territories", and to their statistical offices as "oblast-level statistical administrations" (OSU's).

The local Inspectorates play a much smaller role in the statistical

TABLE A2.2

VERTICAL SUBDIVISIONS OF THE CENTRAL STATISTICAL ADMINISTRATION (TsSU)



* oblast'-level statistical administrations (OSU's)

system than do higher offices. Thus, all industrial and construction enterprises, except those of local subordination, submit their statistical reports directly to the OSU's. The Inspectorates collect agricultural data and various kinds of local information. Aggregated data are passed on to higher levels of the hierarchy and finally to TsSU USSR.

Statistical offices below All-Union level are subject to dual subordination - to governmental authorities at their own territorial level (the Councils of Ministers of Union Republics, the executive committees of oblast' and raion soviets etc.), and to statistical offices at higher levels. It seems that the instructions of TsSU have priority in the event of conflict.

A system of computerised processing of statistical data, known as the Automated System of State Statistics (Avtomatizirovannaya sistema gosudarstvennoi statistiki; ASGS), is now in operation. Every statistical office down to oblast'level accordingly has its own computer centre[4]. The local Inspectorates still use less advanced types of calculating machinery.

The horizontal subdivision of TsSU

The main horizontal subdivisions of TsSU USSR are shown in Table A2.3. The Republican TsSU's have horizontal subdivisions corresponding to almost all of the horizontal subdivisions of TsSU USSR, and the OSU's also have horizontal subdivisions corresponding to most of the horizontal subdivisions of TsSU USSR[5]. A subdivision of an OSU or of a Republican TsSU is subordinate to the corresponding subdivision of the office at the next level above as well as to the director of its own office. It is not clear whether local Inspectorates have any formal subdivisions.

It will be seen from Table A2.3 that horizontal subdivisions are of four types:

TABLE A2.3

HORIZONTAL SUBDIVISIONS OF THE CENTRAL STATISTICAL ADMINISTRATION (TsSU USSR)

Administration for the Balance of the National Economy
Administration of Statistics of Industry
Administration of Statistics of Agriculture
Administration of Statistics of Capital Construction
Administration of Statistics of Material-Technical Supply
Administration for Conduct of the All-Union Population Census
Administration of Statistics of Foreign Countries
Department of Statistics of Transport and Communications
Department of Statistics of Trade
Department of Statistics of Labour and Wages
Department of Statistics of Culture(1)
Department of Statistics of Population, Health-Care and Social Security
Department of Statistics of Finance and Prices
Department of Statistics of Technical Innovation(2)
Department of Statistics of the Environment(2)
Department of Statistics of Residential and Everyday Services
Department of Budget Statistics
Department of Summary (svodnaya) Statistics and of Statistical Methodology
Department for Supervision of the Work of Local Statistical Bodies
Main Administration of Computational Work(3)

Notes:
(1) including statistics of science and education
(2) recently formed departments
(3) responsible for administering the TsSU network of computer centres and mechanised data-processing centres; formerly Glavmekhschet

(a) branch or sectoral departments and administrations responsible for statistics pertaining to a particular branch of the economy (industry, agriculture, construction, trade etc.);

(b) inter-branch or general departments and administrations responsible for statistical work relating to the economy as a whole (balance of the national economy, labour and wages, technical innovation etc.);

(c) departments and administrations dealing with other than purely economic statistics (statistics of culture, population, the environment, budget statistics etc.);

(d) service departments and administrations, such as the Department for Supervision of the Work of Local Statistical Bodies and the Main Administration of Computational Work, which administers the TsSU network of data-processing centres.

To each subdivision of type (a), (b) or (c) there corresponds a "functional subsystem" (funktsionalnaya podsistema; FP) of ASGS, the computer system: for example, the FP "Budget statistics".

The staffing of TsSU

At the beginning of 1977, 20,600 people were employed in the statistical offices of TsSU[6] (Sovershenstvovanie ... 1979 p.17). Of these 35 per cent had some form of higher education and 27 per cent had secondary specialised education (that is, had graduated from tekhnikumy), while 38 per cent lacked post-school education.

An employee of TsSU who has graduated from a tekhnikum, whether from one of the statistical tekhnikumy run by TsSU or from one of the economic tekhnikumy of the Ministry of Higher and Special Secondary Education offering statistics as a specialism, is considered qualified as a "statistician" (statistik). An employee who has graduated from a higher educational institution - a university, an economics institute or the Moscow

Economic-Statistical Institute - is considered qualified as an "economist" (ekonomist).

The terms "statistician" and "economist" are here used to distinguish not fields of specialisation but qualification levels and the corresponding official positions. "Statisticians" are statistical clerks who carry out routine tasks of data recording and processing, while "economists" are responsible for data analysis.

3 Organisation and staffing of the Department of Budget Statistics

The director of OBS TsSU USSR is D.Dunmov, who formerly worked in the Department of Agricultural Statistics of TsSU USSR. The previous director was I.Ya.Matyukha.

The collection of family budgets for the budget survey is, in almost all cases[7], carried out by the OBS's of the OSU's. Computerised data-processing is initially carried out at the computer centres of the OSU's, and aggregated results are submitted to higher-level statistical offices for further processing.

The FBS consists of two separately conducted sub-surveys: the survey of budgets of families of workers and employees, and the survey of budgets of families of collective farmers. Some OSU's collect budgets from both social groups, others collect budgets from only one of the groups, while yet others take no part in the survey at all. The OBS's of those statistical offices concerned with both sub-surveys are divided into two sectors, one for each sub-survey. Thus, in effect, we have not one but two hierarchies of budget survey departments.

The staff of an OBS of an OSU consists of a director, one or several "economists" specialising in budget survey work (ekonomisty po byudzhetam) and a number of "statisticians" specialising in budget survey work (statistiki po byudzhetam). As the main task of "budget statisticians"

is conducting interviews, we shall refer to them as "interviewers". Each "economist" supervises three or four interviewers (Sidlyarenko 1964).

It is of interest to consider the proportion of TsSU statistical staff employed on the FBS. In 1964-5 there were 2,200 interviewers (Krasnoshchekov 1964, Yezhov 1965 p.317). Assuming a ratio of one supervisory "economist" to 3.5 interviewers, this implies about 630 supervisory "economists". Yezhov (1965) reports that about 800 TsSU "economists" in all work on the processing and analysis of FBS data. Therefore about 170 of them work above supervisory level, in Republican TsSU's and TsSU USSR.

It follows that at least 3,000 staff were employed on the FBS in 1964. Between 1964 and 1977 the FBS sample increased by about 1.22 times, from 51,000 to 62,000. If we assume that staff increases proportionally with sample size, then our estimate for the number of staff in 1977 is 3,660, or about 18 per cent of all TsSU statistical staff. Bearing in mind that the staff of local Inspectorates take part in checking the work of FBS interviewers (Chapter C1), we conclude that the proportion of the total working time of TsSU staff devoted to the FBS cannot fall far short of 20 per cent[8].

Notes to Chapter A2

1 We follow the common practice of deriving abbreviations from the Russian names of organisations etc.

2 Apart from the "State statistics" collected and analysed by TsSU, there are "departmental statistics" collected by Ministries and other organisations about their own activity. A considerable amount of statistical research is also conducted by non-TsSU scientific research institutes.

3 The information presented in this chapter has been assembled from a large number of disparate Soviet sources. For fairly useful accounts of the organisation of TsSU see Yezhov (1967, 1968) and Avtomatizirovannaya... (1979, pp.14-17, 118-20, 140-49).

4 TsSU USSR is in fact served by two computer centres: the Main Computer Centre (Glavnyi vychislitelnyi tsentr; GVTs TsSU USSR) and the Central Computer Centre (Tsentralnyi vychislitelnyi tsentr; TsVTs TsSU USSR).

5 Local statistical offices obviously have no Departments for Supervision of the Work of Local Statistical Bodies.

6 A much larger number of people - 175,500 in 1977 - work within the TsSU network of computer centres. This is to a large extent explained by the fact that TsSU computer centres carry out data processing on contract for thousands of other organisations and enterprises as well as for TsSU statistical offices.

7 Raion Inspectorates carry out interviewing in some very sparsely populated areas (Chapter B3).

8 As the FBS is not conducted at all in many oblasti, it accounts for an even higher proportion of staff in those OSU's which do conduct it on any scale. Doroshenko, Head of the OSU of Mogilev oblast' in Belorussia, informs us that 19 out of his staff of 56 are employed on the survey of collective farmers' budgets. Workers' and employees' budgets are not apparently surveyed in Mogilev oblast' (Vsesoyuznoe ... 1959 pp.71-2). In some offices FBS staff must constitute not much less than one-half of the total.

CHAPTER A3

THE HISTORICAL DEVELOPMENT OF MATHEMATICAL STATISTICS AND OF SAMPLING THEORY IN THE WEST, IN TSARIST RUSSIA AND IN THE USSR

1 Introduction

The term "statistics" is used in two distinct senses. It can refer to quantitative information about socio-economic phenomena, or to the study of the general problems of collecting, presenting, analysing, interpreting and using such information[1]. Alternatively, it can refer to the discipline of mathematical statistics, which deals with a particular set of mathematical methods, based on probability theory, for collecting and analysing data of many kinds. Mathematical statistics can be applied both to the collection and analysis of socio-economic statistics and to the collection and analysis of other kinds of data - for example, to the design and analysis of scientific experiments. Conversely, socio-economic statistics can be collected and analysed either with or without the help of mathematical statistics.

This situation also applies to the practice and theory of sampling. Modern probabilistic sampling theory constitutes a branch of mathematical statistics. As such, it can be applied not only to the collection and analysis of socio-economic statistics but also in other fields, such as production quality control. Likewise, socio-economic statistics can be collected with or without the application of modern sampling theory. This theory is not applied both when data are collected by complete enumeration of the population and when non-probabilistic sampling methods of pre-modern origin are used.

The central focus of this thesis is on the application of mathematical

statistics, and of modern sampling theory in particular, to the exercise in socio-economic statistics represented by the Soviet Family Budget Survey. We do not therefore cover the full range of issues either of socio-economic or of mathematical statistics, but consider the intersection between the two.

Contemporary Soviet statistical practice can be properly understood only in a historical perspective. This is much more true of Soviet than of Western statistical practice, because the full application of mathematical statistics to socio-economic statistics continues to be obstructed by factors which have their origin in the Stalinist period. In particular, pre-modern forms of sampling remain in much wider use in the USSR than they do in the West.

This chapter outlines the pattern of historical development of socio-economic statistics, of mathematical statistics, and especially of the interaction of the two. Sections 2 and 3 are devoted to the development of mathematical statistics in general and of its application in socio-economic statistics, while Sections 4 and 5 cover the same ground with respect to sampling in particular. The pattern of development in the West - Western Europe and North America - is dealt with in Sections 2 and 4, the more or less independent pattern of development in Tsarist Russia and then in the USSR in Sections 3 and 5. The main points are summarised in the concluding Section 6.

2 The historical development of mathematical statistics in the West

Socio-economic statistics, in the form of State statistics, can be traced back to the censuses of the ancient riverine civilisations. Mathematical statistics, by contrast, has emerged relatively recently. Its precursor was the school of "Political Arithmetic" founded in seventeenth-century England by Graunt and Petty, who used elementary

probabilistics to study such public issues as the causes of disease (Pearson 1978). Further advances in the analysis of statistical variation were made by such nineteenth-century mathematicians as Poisson, Gauss and Quételet. Mathematical statistics finally emerged in its contemporary form during the period 1890-1940. The most crucial contributions to this process were Galton's early theory of correlation and regression, the "classical" theory of statistical inference of Neyman and Pearson, and the work of Fisher on the analysis of variance.

The socio-economic orientation of Political Arithmetic notwithstanding, the development of mathematical statistics in the nineteenth and early twentieth centuries was primarily motivated by the needs of the physical and especially the biological sciences. The Gaussian study of statistical variation was concerned mainly with the problem of experimental error in applied science. Galton was interested above all in heredity; Gossett ("Student") was led to the t-test of statistical significance by the needs of beer production in the Guinness brewery at Dublin; Fisher was occupied in the design of field experiments between the wars at Rothampstead Agricultural Station. The new discipline was for some time known by the name of "biometrics".

The application of mathematical statistics to socio-economic statistics was delayed by the isolation of State statistics from the new theoretical developments. The officials responsible for State statistics lacked the mathematical training necessary to appreciate the potential value for their work of probabilistic methods, while most of the mathematicians who pioneered these methods appear to have taken no interest in State statistics. Interaction between socio-economic and mathematical statistics was at first promoted only by a few reformers who understood both subjects, such as Professor Bowley, the occupant of the first Chair devoted to statistics in the social sciences (at the LSE). In Britain the gap was

substantially overcome when a more professional Government Statistical Service was built up after World War Two, although the institutional relationship between State statistics and mathematical statistics remains a source of some difficulty even today.

3 The historical development of mathematical statistics in Tsarist Russia and the USSR

There developed in Russia in the late nineteenth century an autonomous tradition of socio-economic statistics, mainly based in the statistical services of the local-government zemstva[3], which was more sensitive to the potential applications of mathematical statistics than the Western State statistics of the time. Such classical Russian statisticians as Chuprov were simultaneously socio-economic and mathematical statisticians.

Soviet statistics of the 1920s in many ways represented a continuation of the zemstvo tradition. The statisticians of TsSU enjoyed both generous State support and a considerable degree of professional autonomy (Wheatcroft 1980). They took a great interest in the work of Western mathematical statisticians, whose methods they developed further and applied to economic analysis. At the same time they criticised Western mathematical statistics for its "empty empiricism" and lack of interest in substantive issues (Yastremskii 1927).

At the end of the 1920s a group of mathematical statisticians associated with the New Economic Policy came under Stalinist attack. Some Stalinist statisticians used this campaign to attempt to discredit mathematical statistics as such, on the grounds that probabilistic methods, while eminently suited to analysis of the anarchy of the capitalist market, were alien to a planned socialist economy. Others argued that, while the "wreckers" had misused mathematical statistics for anti-State purposes, planning could not eliminate all probabilistic phenomena (for example,

the weather) and therefore the correct application of mathematical statistics remained necessary (Smit 1930). Conflict between these two points of view continued throughout the Stalinist period, and the defenders of mathematical statistics were vindicated in the years following Stalin's death.

The dogmatic opponents of mathematical statistics at no time achieved total dominance. Thus, even in the years after World War Two, when they took advantage of the "anti-cosmopolitan" campaign to denounce "enemies of the people ... who propagate bourgeois theories under the slogan of defence of mathematics" (Methodology... 1952), they did not succeed in suppressing statistical methods of quality control in the aviation industry, the practical need for which was realised by the leadership. However, the prolonged influence of the dogmatists has had a powerful impact on Soviet statistics.

First, the interaction of mathematical with socio-economic statistics was terminated. "The possibility of applying the methods of mathematical statistics to the statistical study of social phenomena", and often even "the expediency of mathematical methods of any complexity in statistics", were denied (Nemchinov 1955). Many mathematical statisticians left socio-economic statistics to work in other fields[4]. A "general theory of statistics" emerged which expounded a methodology uninformed by mathematical statistics[5].

A great deal of work on the socio-economic application of mathematical statistics has been carried out since the 1950s. Nevertheless, most of this work has been done in institutes outside TsSU, and has not greatly affected the methods in use within TsSU itself. It is published in such journals as Uchenye zapiski po statistike ("Scholarly Notes on Statistics") and Ekonomika i matematicheskie metody ("Economics and Mathematical Methods") rather than in the TsSU journal Vestnik statistiki ("Statistical Courier").

TsSU has not been among the organisations participating in the series of conferences on the application of mathematical statistics in economics held from 1972 onwards[6].

Moreover, the teaching of mathematical statistics and the teaching of socio-economic statistics remain to a considerable extent isolated from one another in educational institutions. Mathematical statistics rarely occupies a prominent position in the curricula of the institutions which train staff for TsSU (the statistical tekhnikumy, the Moscow Economic-Statistical Institute etc.), while the mathematics faculties of higher educational institutions teach a very abstract and non-applied kind of mathematical statistics.

Second, the dogmatic positions of the Stalin period have not yet been completely overcome. One still occasionally comes across expositions of the view that probabilistic schemas are inapplicable to socio-economic phenomena (Maslov 1971 pp.35-6)[7], or of the view that probabilistic methods contradict the nature of a planned economy (Lipkin 1977). It may well be that positions of this kind will disappear from circulation when the older generation, educated in the Stalin period, leave the scene. However, less explicit attitudes at least partly originating in the earlier dogma may be more persistent.

Soviet approaches to forecasting provide an example of such attitudes (Shenfield 1983a). A.Ya.Boyarskii, the Head of NII TsSU, observes that State statisticians are accustomed to dealing with figures that are (supposedly) uniquely accurate and must therefore overcome a "psychological barrier" before accepting the non-unique results of probabilistic forecasting (Metodologicheskie... 1977 pp.8-9).

A discussion of significance testing by Boyarskii (1980) shows that he himself remains influenced by another tenet of the Stalinist doctrine of statistics - the idea that the function of statistics is to illustrate

theories already known to be true rather than to assess tentative hypotheses. He argues that, even if a test of statistical significance rejects an apparent correlation between the scale of production and productivity of 0.1 as a chance deviation from zero, it is natural for an economist with a theoretical knowledge of economies of scale to consider this a high correlation. From this point of view there is no way that any statistical analysis could ever discredit prior assumptions.

4 The historical development of sampling theory in the West[8]

Most nineteenth-century socio-economic statisticians took the view that only the data of complete censuses could be regarded as "statistics properly speaking". This was an understandable attitude at a time when the only sampling methods practised were widely known to be unreliable.

The most important of these very early sampling methods was "the monographic method", invented by the social reformer LePlay, who from 1829 onwards collected hundreds of detailed "monographs" about the budget and way of life of workers' families (Lazarsfeld 1961). In a monographic survey of a population, an extremely detailed quantitative and qualitative description of a fairly small number of units is obtained. The surveyed units are supposed to be carefully selected by experts in such a way that each "type" of unit in the population is represented in the sample by one unit, or a few units, "typical" of that type (thus the alternative term "typological sampling"). As the critics of monography pointed out,there was no way of verifying the judgement of the sampler regarding typicality. Furthermore, most populations are not comprised of a few known homogeneous "types", and a monographic sample by its very nature cannot reflect heterogeneity within types.

In the 1890s Kiaer, the Director of the new Bureau of Statistics in Norway, developed and used a new form of sampling which he called "the

representative method". Kiaer's method was imitated by Wright, the Director of the US National Bureau of Labour. The method was, in Kiaer's words:

> a partial enquiry in which the observed units are distributed so that their totality forms a miniature of the whole country, a photograph which reproduces the details of the original in its true relative proportions.

To achieve this aim Kiaer used complex multi-stage sample designs incorporating intensive stratification and elements of systematic selection (for example: select males aged 17, 22, 27 ... with names beginning with A, B, C, L, M and N). Strata proportions were determined on the basis of the results of previous censuses, which also served as a means of assessing the representativeness of the sample. The main difference between such "purposive" or "balanced" samples and modern multi-stage sample designs is the absence of random selection within strata.

Probability theory was first systematically applied to sampling in the West by Bowley, who introduced the basic theory of simple random sampling in 1906[9]. The great advantage of random sampling is that, by controlling the probabilities of inclusion of population units in the sample[10], it makes it possible to estimate by means of probability theory the precision of sample estimates in the form of standard errors or confidence limits. The first social survey using probability sampling was conducted by Bowley in Reading in 1912 (Maunder 1977). The theory of probability sampling was extended to stratified random sampling by Neyman and Pearson in 1934.

In the 1930s, large-scale practical experimentation with probability sampling was undertaken by US agencies such as the Bureau of the Census and by the newly formed Indian Statistical Institute. In some countries, such as Sweden, survey sampling before the war continued to rely on "the representative method" (Medin 1983). Probability sampling replaced earlier forms of sampling in State statistics after the war. In market research and opinion polling, however, balanced sampling still remains in use as

"quota sampling". Moreover, sampling theorists in the 1970s have taken renewed interest in the possibility of putting balanced sampling on a sound basis[11].

The relatively late emergence of sampling theory is perhaps the most striking manifestation of the former isolation of socio-economic from mathematical statistics. Both the practical need for sound sampling and the mathematical apparatus for its development already existed in the nineteenth century, but the necessary interaction between the potential suppliers and the potential consumers of sampling theory was lacking.

5 The historical development of sampling theory in Tsarist Russia and the USSR

As in the West, State statistics in nineteenth-century Russia relied mainly on complete censuses. However, various forms of non-probability sampling came into use towards the end of the century.

Most of the studies of peasant household budgets which the statisticians of several zemstva undertook from the 1870s onwards were monographic surveys, based on the selection by one method or another of households supposedly "typical" of different regions (Wheatcroft 1980).

Apparently peculiar to Russia was the form of incomplete enumeration known as "the census method" (tsenzovoi metod)[12]. The "census" (tsenz) here was a register of all those population units considered important enough to be included in the statistics; data were collected on all these units and only on them. Thus, the Tsarist Ministry of Finance maintained in the late nineteenth century a list of "census railway stations" for each type of freight; these lists were used to compile statistics of railway transport (Poplavskii 1927). "Census industry" consisted of enterprises with a minimum size of workforce, depending on the level of mechanisation (Wheatcroft 1981). The rationale of the census method was to use limited

resources to cover the main bulk of the phenomenon of interest. However, there was no way in which the results could be extrapolated to the population as a whole, as the relatively few large units covered were very far from representative of the many small units neglected.

It seems that, unknown to the West at the time, the theory of probability sampling was developed independently in Russia several years in advance of corresponding Western work. The application of probability theory to sampling was first proposed in a paper which Chuprov presented to a congress of scientific research workers as early as 1894. The theory of optimal allocation in stratified random sampling, generally attributed to Neyman's paper of 1934, was set out in a book on sampling theory by Kovalskii, published in Saratov in 1924 (Žarković 1956, 1962)[13].

In the 1920s TsSU felt a great need to develop sampling methods, with practical experimentation often proceeding in advance of theory[14]. The State required statistical information to regulate the economy, but the coverage of all economic units was not necessary to the economic means of regulation used during the New Economic Policy. Nor was complete enumeration practicable given the scattered nature of the NEP economy. Such conditions were very favourable to the development of sampling.

However, although the theory of probability sampling was worked out by some statisticians in the 1920s, probability sampling did not completely replace earlier forms of sampling. The census method, in particular, remained in quite wide use - for example, in the statistics of rail and water freight transport (Poplavskii 1927) and in the study of labour productivity in industry (Akinshina 1966).

The attack on the application of mathematical statistics in socio-economic statistics at the end of the 1920s had an especially deleterious effect on sampling practice and theory, the development of which seems to have been "frozen"[15]. The application of sampling theory was in general

neglected during the Stalin period (Nemchinov 1955). The census method - now renamed "the method of the basic mass" (metod osnovnogo massiva) - continued to be used in such fields as rail transport statistics (Kochetov 1966),and was also applied in the new survey of collective farm markets, which covered only the largest urban centres (Belyaevskii 1962). As we shall see in Chapter A4, the method remains in use even today.

The administration of the command economy set up in the 1930s required the collection of much statistical information on the basis of complete enumeration of economic units. Statistics was in fact identified with national-economic accounting, in which sampling naturally had no place. But, as we shall argue in the next chapter when we consider the position of sampling in the post-Stalin period, even within a command economy sampling could very often substitute for complete statistical reporting, and it was neglected at great cost.

6 Conclusion

The interaction of socio-economic and mathematical statistics has proceeded along very different paths in the West on the one hand and in Russia and the USSR on the other.

In the West the virtual isolation of the two fields from one another which prevailed in the nineteenth century was broken down in the first half of the twentieth. Probability sampling in particular has become a central tool of State statistics, and has on the whole displaced earlier non-probabilistic forms of sampling.

In Russia an independent statistical tradition developed in the few decades before 1917 which proved capable of integrating socio-economic with mathematical statistics and which reached its apogee in the Soviet 1920s. Up to that time Russian and Soviet statistics were somewhat in advance of the West in the field of sampling theory and practice.

However, progress was frozen at the onset of the Stalin period, when the application of mathematical statistics in socio-economic statistics came under sharp attack. An isolation of mathematical from socio-economic statistics was imposed, similar to that which was now disappearing in the West. Sampling was neglected, and the early non-probabilistic forms of sampling remained in wide use. Since Stalin this legacy has been overcome only to a limited extent.

Notes to Chapter A3

1 Methods of collecting, presenting, analysing, interpreting and using socio-economic statistics can be usefully classified into three categories:

(a) the methods based on relatively simple mathematics which mathematical statisticians call "descriptive statistics";

(b) the methods of mathematical statistics, based on probability theory; and

(c) relatively complex mathematical methods not based on probability theory (for example, analysis of indices or non-stochastic programming).

2 In 1906 Professor Bowley addressed the British Association for the Advancement of Science as follows: "... Edgeworth's illustrations in 1885 of the importance of mathematical methods in testing the truth of practical deductions have as yet borne singularly little fruit... It is time that mathematical statistics was brought to bear on the criticism and analysis of existing industrial statistics... Most of our statistics remain untested and their significance not analysed..." (Maunder 1977).

3 The zemstva enjoyed a degree of independence from the central government and were open to external intellectual and political influences.

4 For example, V.S.Nemchinov turned to the design and analysis of agricultural experiments at Bezenchukskaya Experimental Station, developing a computational system based on Chebychev's polynomials (Nemchinov 1946).

5 Reformers in the post-Stalin period have pressed for the methods of the "general theory" to be combined with methods from mathematical statistics. For example, Yuzbashev (1967) criticises the method of "analytical grouping" for ignoring the confounding effect of uncontrolled variables in comparing groups of units, and suggests that it be combined with the analysis of variance.

6 For an account of one such conference, the All-Union Scientific-Technical Conference on the Application of Multivariate Statistical Analysis in Economics and on Production Quality Control, held at Tartu (Estonia) in 1977, see Aivazyan et al (1978).

7 Professor P.P.Maslov was a veteran Soviet statistician, prolific in a number of branches of State statistics. According to his obituary, he was "one of the greatest of contemporary statisticians" (Ryabushkin and Sinclair 1975).

8 For discussions of the early history of sampling, see Stephan (1948), You (1951) and O'Muircheartaigh and Wong (1981), of which the last is the most perceptive.

9 In fact, it was Bortkiewicz who was in 1901 the first in the West to suggest applying probability theory to sampling problems. He recommended the use of Poisson's formula to determine whether differences between census control proportions and sample proportions could have arisen by chance.

10 Simple random sampling provides for equal probabilities of inclusion, which was at first regarded as an essential principle. Random sampling in general provides for known, but not necessarily equal, probabilities of inclusion.

11 It was Royall (1970) who resurrected balanced sampling in the context of a non-Bayesian superpopulation approach. For a discussion of the issue see O'Muircheartaigh and Wong (1981 pp.12-14), who are sceptical as to whether the possible gains from balanced sampling in raising representativeness are likely to outweigh the risks from giving up the ability to estimate precision. Moreover, balanced sampling has been justified only under certain conditions.

12 The census method is sometimes referred to as "concentrated sampling".

13 It is symptomatic of the subsequent fate of sampling in the USSR that Soviet authors do not make reference to Kovalskii's work in discussing the origins of sampling. It was the Yugoslav statistician Zarković who rediscovered Kovalskii's book in the Lenin Library in Moscow.

14 Conferences of statisticians instructed the Methodological Section of TsSU to develop the theory of methods found necessary in practice, such as cluster sampling (Zarković 1956). For an account of the sampling methods used in Soviet sociology in the 1920s, see Sheregi (1978).

15 For a possible example of the "freezing" of sampling methodology, see Chapter A4 Section 4(e).

CHAPTER A4

THE POSITION AND PRACTICE OF SAMPLING IN CONTEMPORARY SOVIET STATISTICS

1 Introduction

In this chapter we do not intend to provide a full or balanced survey of the position and practice of sampling in contemporary Soviet statistics, but only to make some general observations and to bring together those points which will facilitate understanding of the way the FBS is conducted.

We start in Section 2 by reviewing Soviet terminology concerning "forms of statistical observation" and explaining its significance. In Section 3 we describe the subsidiary position of sampling within Soviet statistics, and draw attention to some common Soviet attitudes towards sampling which contribute to its continuing relative neglect. Section 4 deals with the survival of non-probabilistic forms of sampling in Soviet practice: monographic observation, the method of the basic mass, directed sampling and the questionnaire method.

In Section 5 we discuss four deficiencies[1] characteristic of much Soviet probability sampling, and from which the sample design of the FBS also suffers:

(a) The "intuitive" determination of sample size without sample-theoretic calculation and with a tendency to prefer unnecessarily large samples (for corresponding discussion of the size of the FBS sample, see Chapter B2);

(b) The neglect of the economics of sampling, which is associated both with the tendency towards excessive sample size and with a tendency

to use inefficient sample designs;

(c) The practice of starting in the middle of the first interval in systematic sampling, instead of selecting the starting-point at random, which gives rise to a "tail-cutting bias" (analysed with respect to the FBS in Chapter B1); and

(d) The practice of carrying out representativeness checks by comparing population with sample means alone (considered in the case of the FBS in Chapter B6).

We sum up the key points in Section 6.

2 "Forms of statistical observation" in Soviet practice[2]

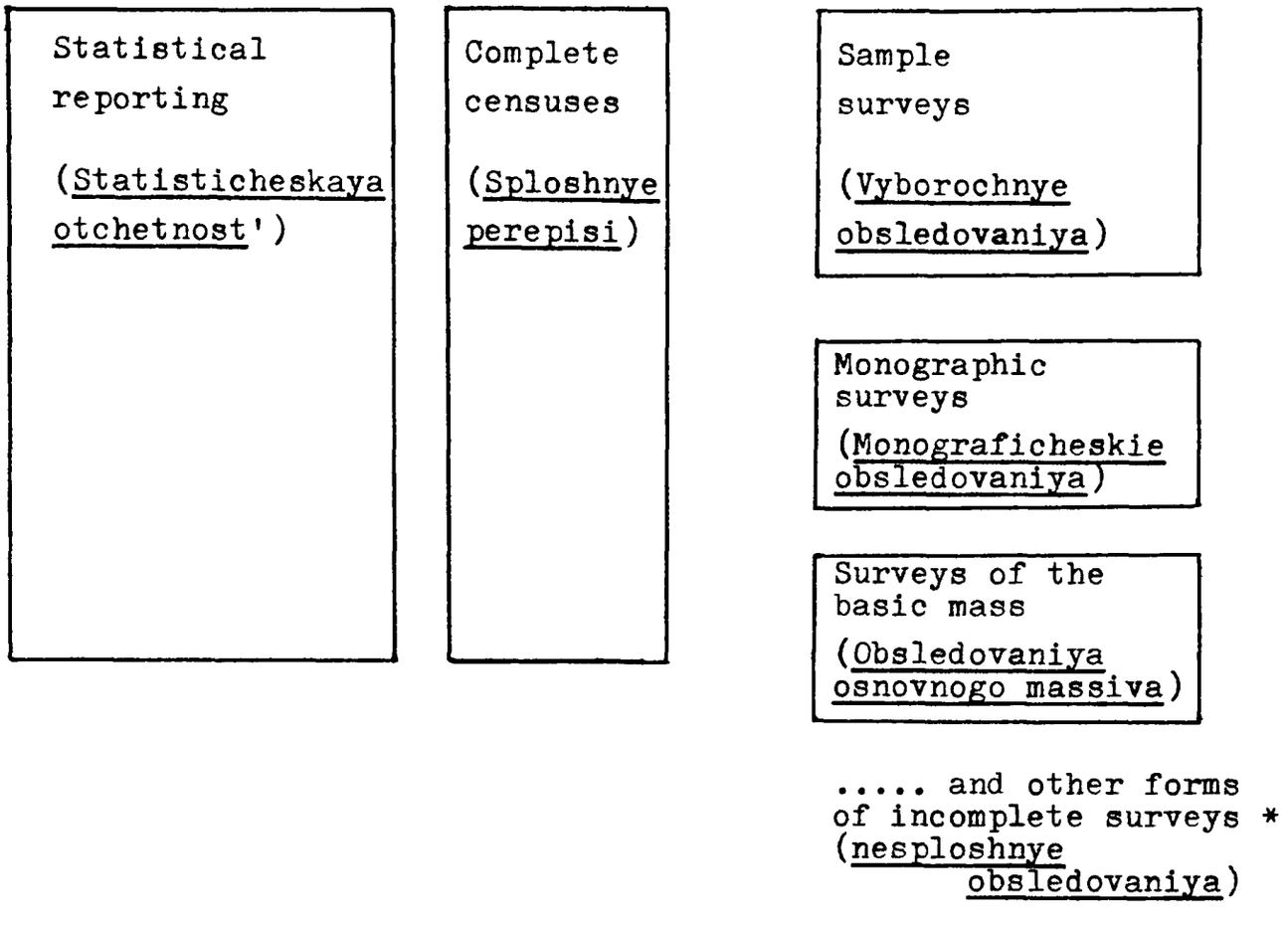
The terminology used in the USSR to classify "forms of statistical observation" is in itself very revealing of the nature of Soviet statistical practice. The key terms with which we are concerned and the relationships among them are shown in Table A4.1.

The three columns represent the three main categories of forms of statistical observation: the comprehensive and periodic statistical reporting of all economic enterprises and other establishments and organisations (such as hospitals and passport offices) to TsSU (otchetnost'); censuses (peripisi)[3]; and surveys based on incomplete enumeration of the population. Censuses and surveys may collect supplementary data from enterprises and establishments which also submit otchetnost' (for example, the census of uninstalled industrial equipment), or they may collect data from individuals or families (as in the population census or the FBS).

Forms of statistical observation are also subject to two overlapping dichotomies. One of these is basically organisational in nature: the division between reporting, the administrative apparatus of which is in continuous operation, and all other forms of observation, which have to be "specially organised". The other dichotomy is a methodological one: both

TABLE A4.1

FORMS OF STATISTICAL OBSERVATION IN SOVIET PRACTICE



* For example, surveys by "the questionnaire method" (anketnyi metod)

reporting and censuses are based on complete enumeration of the population, while surveys are based on incomplete enumeration.

The order in which forms of observation are always listed reflects their respective positions within the statistical system. The central bulk of statistical work is concerned with reporting. Censuses are regarded as a supplement to reporting, and incomplete enumeration is regarded as a supplement to complete enumeration.

The subdivisions within the category "incomplete surveys" arise from the persistence of various forms of non-probability sampling (monographic survey, method of the basic mass etc.) in Soviet practice alongside probability sampling. Thus the distinction is often made between sample surveys proper (vyborochnye), in the sense of sample surveys designed in accordance with the modern theory of probability sampling, and incomplete surveys in general (nesploshnye), which include surveys based on either probability or non-probability sampling[4].

3 The position of sampling within Soviet statistics

The status of sampling in the post-Stalin period is described by V.Ye.Ovsiyenko (1966) as follows:

> Sampling no longer has opponents ... so long as we restrict the discussion to theory. Doubts often arise when we come to practice. (There is) uncertainty regarding, and distrust of, the results of sample observation. This is why sample surveys have not achieved as wide an application in our State statistics as they can and must achieve. It is apparently also for this reason that TsSU has not published the results of a number of large-scale sample surveys it has conducted in the fields of population, industry etc.

The assumption is commonly made that, wherever complete enumeration is feasible, it should be carried out, because it is "always richer and more precise" (Volkov 1971). Sampling is usually justified as a necessary expedient for cases in which the ideal of complete enumeration is in

practice unattainable (for example, Aganbegyan et al 1961 pp.137, 154). Soviet authors less often realise that sampling may be preferable even where complete enumeration is feasible: sampling is more cost-effective and may even be more precise if it improves data quality beyond the margin necessary to compensate for sampling error[5].

Sampling inevitably plays a lesser part in a statistical system which is responsible for monitoring the performance of all individual productive units as well as for providing and analysing information about statistical aggregates than it can play in a more autonomous statistical system fulfilling only statistical functions proper[6]. However, a recurrent theme in Soviet discussions is that there is enormous scope for replacing complete reporting by sample surveys even in the absence of reform of the existing economic system. Many proposals have been made for replacing complete enumeration by sampling in particular areas. Some such proposals have eventually been adopted - for example, the last two population censuses have used sampling for parts of their programmes - but many others have been ignored. The subordinate role of sampling is the result not only of the technical requirements of a command economy but also of ignorance and distrust of sampling[7].

The predominant position occupied by complete reporting in the Soviet statistical system has various effects on the nature of sampling work that is carried on. Some sampling practices are directly derivative of complete reporting, such as sample processing of the data of complete reporting (when it is impracticable to process all the data) or sample checking of their reliability[7]. The organisation of complete reporting on the economic branch principle is carried over into many of the sample surveys conducted by TsSU, including the FBS (Chapter B10). The consequent discrepancy between the sampling unit (the individual worker) and the unit of observation (the family) is the source of severe bias in the FBS

sample (Chapter B1).

4 The survival of non-probabilistic forms of sampling in Soviet practice

Non-probabilistic sampling continues to play a substantial role in Soviet statistics. The extent to which such forms of sampling have survived from earlier times in Western and in Soviet practice is compared in Table A4.2.

TsSU conducts what it calls "monographic surveys" quite frequently. One example is the survey in 1968 of the use made of qualified staff in 200 industrial enterprises (O vazhneishikh... 1968). The term appears to be used loosely to refer to any survey in which "expert selection" partially or fully replaces random sampling. Accounts of the sampling methods used may show the influence of ideas related to probability sampling and the method of the basic mass together with that of the ideas of traditional typological monography. The susceptibility of monographic surveys to bias is sometimes defended on the grounds that they do not aim at representativeness (Yezhov 1967 pp.171-2). However, conclusions are drawn from the results of such surveys which are valid only if the samples are representative.

The method of the basic mass is still used in a number of areas. For example, it is found occasionally in sociological research (Sheregi 1977). Its most significant application is in the survey of collective farm markets, the basic methodology of which has not changed since the 1930s. Although there were in 1957 collective farm markets in 4,599 localities (2,812 towns and settlements of urban type and 1,787 rural raion centres), the survey covered only 264 localities, all of them large and medium-sized towns, accounting for 58 per cent of the urban population. The markets covered by the survey are therefore in no way representative of the population of markets as a whole[8].

TABLE A4.2 - OCCURRENCE OVER TIME OF DIFFERENT FORMS OF SAMPLING IN THE WEST AND IN RUSSIA/USSR

Form of sampling	Occurrence in the West	Occurrence in Russia / the USSR
Monographic or typological sampling	Developed by LePlay from 1829 onwards. Superseded early in 20th century.	Used by zemstvo statisticians from the 1870s onwards. Became less common but still occurs.
Concentrated sampling	-	Used by Government Departments from late 19th century and by TsSU in 1920s as "the census method" (tsenzovoi metod). Still occurs as "method of the basic mass" (metod osnovnogo massiva).
Purposive or balanced sampling	Developed by Kiaer in 1890s as "the representative method". Superseded in State statistics by about 1940. Still occurs as "quota sampling" in market research.	Used by TsSU in 1920s. Still occurs in sociological research as "directed sampling" (napravlyënnyi otbor).
Probability sampling	Developed by Bortkiewicz and Bowley in 1900s. Basic development complete by 1934. Dominant form of sampling from 1940s onwards.	Developed by Chuprov in 1890s. Basic development completed in 1920s. Integration into practice "frozen" from 1929 onwards. Still co-exists with other forms of sampling.

A form of non-probability sampling in use by Soviet sociologists is "the questionnaire method" (anketnyi metod). Unlike monography or the method of the basic mass, the questionnaire method does not seem to have its origins in nineteenth-century practice. It is simply a matter of "people with insufficient awareness of and little competence in sampling theory" using "the most accessible methods of information collection" (Shlyapentokh 1976). Questionnaires are distributed on a very large scale among potential respondents with the expectation of a low response rate. The questionnaire may be published in a magazine, given out in trains or left in a pile in a shop. Those who in effect volunteer to take part in the survey are rarely representative of the target population[9].

Although many Soviet statisticians are very critical of non-probability sampling, there are also those who dispute "the view of foreign and some Soviet authors" that only random or systematic sampling is scientific (Maslov 1971 pp.35-6). A social statistician writes:

> In recent years the point of view has become widespread in our sociology that it is possible "to refrain from random modes of selection in forming a sample, giving preference to methods of typological and directed selection and to quota sampling". This statement is justified within certain limits.
>
> (Zhabskii 1983, quoting Standartizatsiya... 1981 p.158)

The advantages of balanced sampling have in recent years, as we have seen in Chapter A3, been reconsidered also by Western sampling theorists. However, while the Western discussions have been conducted on the basis of modern theory, the Soviet debate reflects the continuing influence of outdated philosophical conceptions.

5 Some common deficiencies of Soviet probability sampling

(a) The "intuitive" determination of excessively large sample sizes

A number of critics of Soviet sampling practice observe that

unnecessarily large samples are commonly used (Nikolayeva 1974, Ignatovich 1975, Shlyapentokh 1976). They attribute this tendency to the fact that sample sizes are most often determined "intuitively" without sampling-theoretic calculation (Safronova 1968). The people who decide the sample size do not realise that quite small samples can provide results of precision adequate for practical purposes[10]. Moreover, being unfamiliar with sampling theory, they make the error of judging the effectiveness of a sample on the basis of the sampling fraction, when in fact sampling error depends primarily on the absolute sample size (provided that the sampling fraction is small compared to one, which it usually is)[11]. Safronova (1968) reports that the sampling fractions usually preferred are in the region of 10-20 per cent, while Shlyapentokh (1976) notes that "some sociologists wrongly consider that a sampling fraction of 10 per cent or more automatically guarantees representativeness". Given any but a very small population, sampling fractions of this order of magnitude generate extremely large samples[12].

Even a sample which is extremely large in absolute terms may be regarded with suspicion if the sampling fraction is small. For example, at the conference of statisticians in 1968 the prominent demographer Professor B.Ts.Urlanis criticised the plans for a survey of the incomes, family composition and living conditions of 250,000 families:

> I consider this unconvincing. We have 70 million families, and 250,000 is only one-third of one per cent. This sampling fraction is too small... (Vsesoyuznoe... 1969 p.224)

P.G.Pod'yachikh, Head of the Population Census Administration of TsSU, said in reply:

> Some scholars consider that the survey data will be unrepresentative because the sample is supposedly small - about 0.5 per cent of all families. However, one cannot call a sample of almost a million people "small"... (p.276) [13]

The assumption that the sampling fraction is the best indicator of

the effectiveness of a sample may be based on a feeling that a sample can be trusted only to the extent that it comes close to complete enumeration.

The following conversation between this writer (SDS) and N.Rimashevskaya and A.Shevyakov of the Central Economic-Mathematical Institute demonstrates both the attitude taken by a Soviet economist of the older generation towards sample size and sampling fraction and her ideologically conditioned rejection of the applicability of mathematical statistics to the study of socio-economic phenomena:

SDS shows Rimashevskaya and Shevyakov a graph of how the sample size of the Soviet FBS has changed over time.

Rimashevskaya: But that is meaningless. The population of the USSR has grown enormously over that period.

SDS: I have another graph for the sampling fraction. But the absolute sample size matters as well. The sampling error depends on it.

SDS writes down some formulas for sampling error. Rimashevskaya starts to speak, but Shevyakov attempts to forestall her.

Shevyakov: Of course, yes, the sampling error...

Rimashevskaya: Well, maybe this is just my personal opinion, but all the same I think these formulas apply to quality control in industry but they do not apply to social processes. Social processes are too complex and heterogeneous to be handled by such formulas.

(Shenfield 1982b)

A further factor which doubtless contributes to the tendency towards unnecessarily large samples is the generous funding of statistical work in the USSR. The absence of tight financial constraints on statisticians may explain why the economics of sampling is largely neglected[14].

(b) Inefficient sample design

The neglect of the economics of sampling has the consequence that samples tend to be not only unnecessarily large but also inefficiently designed, in terms of the precision achieved per unit cost (Andreichenko 1975, Shlyapentokh 1976). Provided that sampling-theoretic calculations are made at all, inefficiency of design leads to larger sample size, as a larger sample is needed to achieve a given precision with an inefficient design than with an efficient one.

An example is provided by the design of a sample survey of wages in an unidentified branch of industry conducted in 1971 by the Department of Statistics of Labour and Wages of TsSU USSR. L.A.Nikolayeva (1974), a staff member of NII TsSU, compared several possible stratified cluster designs for the survey, and recommended as the most efficient a three-stage design involving the selection of enterprises, workshops and workers. However, the TsSU Department decided to use a simpler two-stage design which left out the workshop level. Nikolayeva shows that the design recommended by her would, by comparison with the design actually adopted, have reduced sample size by almost three-quarters and cost by almost one-half for the same precision[15].

(c) Incorrect conduct of systematic sampling

It is often convenient to replace random sampling proper by systematic sampling from a list of population units. If the units on the list are numbered consecutively from 1 to N and a sample of size n is to be selected, the list is divided into n intervals of length (N/n) and one unit is selected from each interval[16]. The units selected are those numbered k, (k+i), (k+2i)... , where i is the interval length and k is the "starting-point" ($k \leqslant i$).

If the population units are listed in a way which is random with

respect to all variables of interest, as (usually) in the alphabetical listing of names, then systematic sampling is equivalent to random sampling. It may, however, be decided to list the population units in ascending or descending order of some "ordering variable" of importance to the subject under study. The sample is then guaranteed to have approximately the same distribution on the ordering variable as the population, provided certain conditions are met. Such "ordered systematic sampling" (ranzhirovannyi mekhanicheskii otbor) is frequently used in the USSR. As we shall see in Chapter B1, it plays a central role in the design of the FBS sample.

Ordered systematic sampling can be regarded as a satisfactory substitute for random sampling proper, but only under definite conditions. One of these is that the starting-point must be selected from within the first interval at random, so that all units have a chance of inclusion in the sample. The use of a fixed starting-point can lead to significant bias. Without a random start, in fact, systematic sampling is no longer a form of probability sampling.

In Soviet practice systematic sampling as a rule proceeds from a fixed starting-point, usually from the middle of the first interval[17]. Although a mid-interval start leads to less bias than any other fixed start, it can lead to various significant biases, of which the most serious is the exclusion from all possible samples of the extreme ends of the population distribution of the ordering variable, which we shall call "the tail-cutting bias".

Let us take as an example the sample checks of the accounting of milk production on collective farms conducted by TsSU (Merlinov 1966). Within each oblast' (or krai), four raiony are selected in such a way as to cover the basic zones characterised by different dairy conditions, and two collective farms are selected in each of the selected raiony. Collective farms are selected systematically from lists of the collective farms in

each raion, ordered by number of cows on the collective farm. As a mid-interval start is used, it is always those collective farms placed one-third and two-thirds of the way down each list which are selected. Thereby collective farms with very many or very few cows, relatively for their raion, are excluded from the sample check.

Mid-interval starts are also used in the FBS sample design. Their effects are examined in Chapter B1.

Mid-interval starts are not only generally used in practice but are also recommended in almost all texts offering guidance on sampling, including those written by statisticians who seem on the whole to understand sampling theory, such as V.Ye.Ovsiyenko (1966):

> If units are ordered on a variable under study, then selection should begin from the middle of the first interval, in order to avoid bias... High precision of sampling is guaranteed by the inclusion in the sample of more or less typical, average representatives of the parts into which the population is in essence divided by systematic selection.

One sees in this passage the continuing influence of the monographic concept of sample "typicality" and of the associated neglect of the problem of ensuring representativeness with respect to the population distribution as a whole. We have found only one source which explains the need for a random start in systematic sampling; the author is P.O.Kenkmann (1968), a student in the History Faculty of Tartu University (Estonia)[18].

(d) Inadequate checks of sample representativeness

In Soviet practice it is usual to check the representativeness of a sample by comparing sample with population means on one or more variables of importance. If the discrepancy between the two means falls within a given margin, generally expressed in percentage terms, the sample is regarded as sufficiently representative. Larger discrepancies are "corrected" by replacing sample units with extreme values on the check variable by other

population units. The replacement units may be selected by a variety of methods: selection may be random, or (in systematic sampling) units adjacent to the excluded units on the list of population units may be chosen (Ovsiyenko 1966).

For an account of representativeness checks of this kind in the field of agricultural statistics, see Samoilov (1966). We shall consider the practice as applied to the FBS in Chapter B6.

Critics of Soviet sampling practice agree that sample and population means should be compared[19], but point to various shortcomings of the method and argue that one should not rely solely on it; it is essential also to determine sampling errors (Ovsiyenko 1966, Safronova 1968). The checks used usually cover means only, and so do not reveal biases in sample distributions which leave means unaffected. The percentage criteria on which discrepancies are assessed have no justification in sampling theory, and are apparently set "intuitively". Ovsiyenko (1966) also draws attention to the problem that the population data used in the checks may be outdated.

Like the use of mid-interval starts in systematic sampling, reliance on the comparison of population and sample means in representativeness checks is suggestive of the continuing influence of the monographic concept of typicality. The practice might also be attributed to the "freeze" imposed on the development of Soviet sampling at the end of the 1920s, to which reference has been made in Chapter A3[20]. It may be noted that, shortly before the onset of the "freeze", the TsSU journal Vestnik statistiki published a long report of the pioneering work on sampling conducted by the Italian statisticians Gini and Galvani (Gini 1929). Gini and Galvani had experimented with the selection of samples from the data of the Italian population census of 1921 (Gini 1928). By trial and error they selected a sample of 29 out of the 214 districts (circondari) of Italy such that the average sample values of seven important variables (birth rate, death rate

etc.) were all close to the corresponding population averages for the country as a whole. However, large differences were found between this sample and the population when the averages of other variables were compared, as well as when statistics other than the average were compared on the seven control variables. The theoretical explanation of these biases, and of the dangers of relying on a few comparisons of sample and population means, was provided in the further development of sampling theory by Neyman (1934). But by the mid-1930s Soviet statistics was no longer open to new ideas about sampling.

Representativeness checks by comparison of sample and population means are facilitated by the availability of a great deal of data about population means from complete reporting. Reliance on such checks could be considered another effect of the predominant position of complete reporting in the Soviet statistical system (see Section 3). As complete reporting is compiled by the accumulation of totals or averages at successive hierarchical levels of the statistical system (a process known as "decentralised summarisation"), it does not contain information on the distribution of variables other than their means; such information has to be collected separately by means of specially organised observation. Thus it is practicable to compare population and sample means in many more cases than those in which it is practicable to compare population and sample values of other distributional statistics. The lack of available data on population variability also makes it more difficult to calculate sampling errors. There are then both conceptual and practical obstacles to improving the practice of representativeness checks.

6 Conclusion

We have seen that sampling occupies a position within the Soviet statistical system subsidiary to that of complete enumeration for both

administrative and attitudinal reasons. It continues to be commonly regarded with distrust as an inferior form of statistical observation. Ignorance and suspicion of probabilistic sampling theory remain widespread, with the result that earlier forms of non-probability sampling are still in quite wide use, while the practice of probability sampling often suffers from serious deficiencies, including unnecessarily large samples, inefficient sample designs, errors in the conduct of systematic sampling, and inadequate representativeness checks. These deficiencies also affect the sample design of the FBS.

Notes to Chapter A4

1 To counteract our inevitably one-sided focus on the deficiencies of Soviet practice, we should point out that quite a few well-designed sample surveys are carried out in the USSR - for example, the social surveys conducted by the Institute of the Economics and Organisation of Industrial Production (under the Siberian Division of the Academy of Sciences) and by the Sampling Laboratory of NII TsSU USSR. It is also of course true that sample surveys in the West are of very variable quality.

2 In this Section we draw on the discussions of "forms of statistical observation" to be found in any Soviet textbook on socio-economic statistics - for example, Obshchaya... (1980). See also Shenfield (1982a).

3 It is convenient to be able to distinguish between "censuses", as based on complete enumeration, and "surveys", as based on incomplete enumeration. Unfortunately, Soviet sources do not consistently make this distinction: "censuses" may be incomplete (nesploshnye perepisi) and "surveys" may be complete (sploshnye obsledovaniya).

4 But often too the distinction is not made, all incomplete surveys being called "sample surveys". The word "vyborochnyi" is therefore ambiguous, and its exact meaning must be induced from the context. One author may choose to call a survey based on non-probability sampling, such as the survey of collective farm markets, "vyborochnoe" in order to imply that the survey is just as good as a sample survey proper. Another author may call the same survey "nesploshnoe" as a way of drawing attention to the fact that it is not a "real" sample survey. There is also controversy concerning which of the two words should be used to refer to the FBS, the design of which is influenced by sampling theory but not fully in accordance with it. This controversy is really about the seriousness of the deficiencies of the FBS sample; to deny that the FBS is "vyborochnoe" is to underline its lack of representativeness (see Chapter B12).

5 "The ideal basis for revealing the regularities of effective demand would be complete (sploshnye) data on the incomes and structure of expenditures of each family" (Frenkel and Lakhman 1966). Ovsiyenko (1966) is one of those who demonstrate awareness of the full advantages of sampling.

6 For a discussion of some of the implications for data reliability of the difference between the two types of statistical system, see Shenfield (1983c).

7 This issue is discussed at greater length in Shenfield (1982a).

8 For a full description of the methodology of the collective farm market survey, see Pletneva (1966). An assessment of the representativeness of the survey is provided by Belyaevskii (1962).

9 For an example of a survey by the questionnaire method, in which questionnaires were given out to visitors at the registration windows of a polyclinic, see Orlean et al (1982). For a discussion of the biases entailed by the method, see V statisticheskoi... (1961).

10 See the report of the lecture by A.Ya.Pishchanok on market research at the Leningrad House of Scholars (Ignatovich 1975).

11 Sampling error can in general be expressed as the product of a term independent of population size or of sampling fraction, which is the sampling error for an infinite population, and the "finite population correction".

The finite population correction is $\sqrt{1 - \frac{n}{N}}$

(n = sample size, N = population size), which approximates to 1 for $n \ll N$.

12 Sheregi (1977) analysed the content of all the 500 reports of sociological surveys published in the USSR in 1970-73. Only 66 per cent of the reports gave the sample size. Sampling fractions in this group of surveys were distributed as follows:

Sampling fraction	less than 10%	24% of the surveys
"	10 - 20%	37% " "
"	23 - 30%	13% " "
"	30 - 50%	18% " "

The remaining 8 per cent of the surveys had sampling fractions in excess of 50 per cent. One survey, for example, had a sample of 319 taken out of a population of 515.

13 Nevertheless, the survey in question was unrepresentative, not of course because of its sample size but because it was, like the FBS, organised on the branch principle (see Chapter B1).

14 Both Nikolayeva (1974) and Shlyapentokh (1976) mention that the economics of sampling is neglected. The question of funding is not discussed in the sources, but Konstantin Miroshnik, who took part in the organisation of social surveys in the Ukraine, links the use of very large samples to the looseness of financial constraints (oral communication).

15 The cost saving is not as great as the reduction in sample size because more efficient and therefore more complex designs are more expensive to implement than simpler designs of the same sample size. Officials unfamiliar with sampling theory may have been misled by this fact to prefer a simpler design, assuming that simpler designs are more cost-effective as well as cheaper per unit sample size.

16 To avoid complications of negligible importance, we are assuming that N is exactly divisible by n. For a fuller account of systematic sampling, see Kish (1965), Chapter 4.

17 According to Kenkmann (1968), those who design sociological surveys often take the beginning of the first interval as the starting-point of systematic sampling. On the other hand, one also finds examples of properly conducted systematic sampling, as in the sample survey which formed part of the population census in 1979 (Raikh and Volkov 1980).

18 Matyukha (1960) and Ananyeva (1964) also recommend mid-interval starts. Some authors mention mid-interval and random starts as alternative options without expressing a preference for one over the other (Kamyshev 1972, Venetskii and Kildishev 1975 p.220).

19 The method is "a definite achievement of Soviet statistics" (Ovsiyenko 1966).

20 This historical explanation is advanced only by way of a tentative hypothesis.

PART B

THE SAMPLE DESIGN OF THE FAMILY BUDGET SURVEY

CHAPTER B1

GENERAL PRINCIPLES OF THE SAMPLE DESIGN

1 Introduction

The selection of families for the FBS is carried out as a large-scale exercise only on the infrequent occasions, as in 1952 and in 1969, when the survey undergoes substantial expansion and reorganisation. Sample selection at other times amounts mainly to the piecemeal replacement of individual families which for one reason or another drop out of the survey. There is no regular rotation of the sample; efforts are made to retain participating families in the survey for as long as possible. In this chapter we describe and assess the basic methods used to select families for the sample, both during exercises of sample reconstruction and in replacing drop-outs.

The FBS is organised separately and on different principles for families of workers and employees and for families of collective farmers[1]. We describe the methods used for these two subsamples in Sections 2 and 3 respectively. We proceed to a critical assessment of these methods in Section 4. Three types of sampling bias are inherent in the sample design, apart from the biases arising from the incomplete and uneven coverage of the population which are assessed in Chapters B2 - B4. The three types of bias - "tail-cutting bias", "multi-worker bias" and biases due to excessive period of participation - are discussed in Sections 5, 6 and 7 respectively. We summarise our conclusions in Section 8.

One aspect of the significance of the FBS for Soviet survey sampling more generally is the use made of its sample in the conduct of other surveys. This practice is the subject of Appendix 1 to this chapter.

Before 1952 the collective farmer sample was designed on different lines than thereafter. The former design is discussed in Appendix 2 to this chapter.

2 Selection of families of workers and employees

Budget surveys are conducted in most countries nowadays on the territorial principle. A geographical framework is used to select residential addresses and the households resident at those addresses are asked to participate in the survey. In the USSR,however, the traditional use of the branch principle has never been superseded. Within each branch of the national economy and of industry in which workers and employees are selected for the survey, a certain number of workplaces - industrial enterprises, non-industrial institutions and establishments - are selected, and at each such workplace a certain number of workers and/or employees are selected. The families of which those workers and employees are members then become participants in the survey. We shall see the difficulties entailed by this approach at many points in this thesis.

The first step in a sample construction exercise is the setting by TsSU USSR of sample quotas ("control figures") for each branch by Union Republic (Karapetyan 1980 p.242). Karapetyan urges that in future quotas for geographical regions also be set, no doubt for the purpose of ensuring greater territorial representativeness. However, at present Republican TsSU's are responsible for allocating their quotas among the oblast'-level statistical administrations subordinate to them, and the latter select workplaces, and within them workers and employees, for the survey[2].

Official and textbook accounts always affirm that allocation of the sample among both branches and territorial units is carried out in proportion to the number of workers and employees in the corresponding populations, as shown by statistical report data on wages and labour (Matyukha, Postnikov

and Samoilov 1958; Matyukha 1966, 1967; Kildishev et al 1980; Posobie... 1980). In Chapters B2-B4 we shall see that proportionality of allocation is very far from applying in practice. It seems that proportionality is regarded as a desirable ideal, and the ways in which the real sample falls short of it as embarrassing facts to be ritually denied.

Let us now suppose that the number of workers and of employees to be selected for the sample in a particular branch of the economy - say, engineering and metalworking - within a particular oblast' (krai, ASSR, small Republic) - let us call it oblast' X - has somehow been decided upon. In how many workplaces (enterprises, establishments) will these workers and employees be surveyed, and how will the workplaces be selected?

Let us call the group of budgets collected by a single interviewer an "interviewer set", and the group of budgets collected at a single workplace a "workplace set". The survey is organised so that any one interviewer collects budgets at only one workplace, which is convenient for her because it means she need approach only one bookkeeping office for wage records and because it usually ensures that her families live in roughly the same neighbourhood. Thus a workplace set cannot be smaller than an interviewer set. The two sets are usually identical, with one interviewer attached to each workplace surveyed, but at a very large workplace the workplace set may consist of two or more interviewer sets.

Table B1.1 sets out information on the size of interviewer sets given in various sources. The standard size seems to be in the range of 20-26 budgets; for example, Vladykin (1955) reports 255 workers surveyed at ten enterprises in Kuybyshev oblast'. However, it seems that interviewer sets can sometimes be rather smaller than 20. We shall assume standard inverviewer sets of size 25.

Panina (1974) provides an example of the workplace set comprising more than one interviewer set. At the Kolomenskii Diesel-Locomotive Construction

TABLE B1.1

REFERENCES TO SIZES OF INTERVIEWER SETS

Year	Place[1]	Sizes of interviewer sets (Number of family budgets) Workers and/or employees[2]	Collective farmers	Source(s)
Up to 1951			18[3]	Postnikov (1953, 1966), Krylov (1957)
1952		20 w+e[4]		Krylov (1957)
1953			18-23[5]	Postnikov (1953)
1955		20 (minimum)	12, 18 or 23[5]	Grankov (1955)
1955	Kuybyshev oblast'	25 w		Vladykin (1955)
1957			18 or 23-24[5]	Krylov (1957)
1960-61	Tadzhikistan	20,22,26 w[6]		NKhTadzhSSR v 1960g (Dushanbe 1961), NKhTadzhSSR v 1961g (Dushanbe 1962)
1964		20 (average)	20 or 25	Ananyeva (1964)
1965	(	15-20 w		Yezhov (1965)[7]
	(	20-23		
	(	23.2		
1966			about 25	Matyukha (1966)
1967		20-25		Matyukha (1967)
1968	Odessa oblast'	22 w[8]		Klebanov (1968)
1968	unidentified oblast'	14,16,17 w[6]		Venetskii and Matyukha (1968)
1980		23-25 w or e		Posobie... (1980 p.371)

Notes to Table B1.1

1 Where place is not indicated, reference is to the USSR as a whole.

2 Abbreviations used are "w" for workers, "e" for employees and "w+e" for workers and employees included in a single set. Figures between the two columns refer to workers, employees and collective farmers without distinction (or taken together).

3 Up to 1951 each interviewer collected two sets of budgets: 18 at each of two collective farms in the same raion (see Appendix A to this chapter).

4 Krylov goes through an illustrative exercise relating to 1952 in which sixteen workers and four employees are selected at each enterprise.

5 These references are accounted for most plausibly by supposing that the old pre-1952 sample design (see note 3) was still used in some places in the mid-1950's. Grankov (1955) explains that different norms (for the size of budget sets) are laid down in different oblasti.

6 The reference gives a breakdown of the sample for this area by branch of industry. These figures are the sizes of those branch subsamples which we presume must have been drawn from a single enterprise.

7 Here, the figure of 23.2, calculated by dividing total sample size for the USSR by the number of interviewers (both these being given by Yezhov), is probably the most reliable. It is somewhat inconsistent with the statement by Yezhov (p.317) that 15-20 workers are surveyed in each enterprise, unless we assume that a few employees are also usually surveyed at each enterprise (see note 4). Nor can 23 be the true upper limit for an interviewer set if average set size is 23.2.

8 These are workers on a State farm.

Plant named after Kuybyshev in Moscow, 125 workers and employees participate in the FBS, so that five interviewers are needed to cover this single giant enterprise. This does not appear to be an exceptional case: the number of participants in the FBS at the Productive Association of Kolomenskii raion in Moscow oblast' is also reported as 125, the number of participants at the Glukhovskii Cotton Combine named after Lenin in the town of Noginsk (near Moscow) as 110 (Sobranie... 1982).

We are in a position to estimate roughly the proportion of workplaces in the sample with more than one interviewer. Parfenova tells us that, when the sample was expanded in 1969, 10,500 workers and employees were selected as new participants at 357 newly selected workplaces, implying an average workplace set of 29.4 budgets (Soveshchanie... 1969). For 1965 we have a reliable figure for the average size of an interviewer set: 23.2 budgets (Table B1.1, note 7), so on average a workplace set corresponds to 1.27 interviewer sets. The first interviewer set at each workplace accounts for 8,282 budgets, leaving 2,218 budgets for about 96 additional interviewer sets. As there are about 453 interviewer sets in all, about 21 per cent of all interviewer sets are not first sets. Were there a maximum of two interviewer sets at any one workplace, the proportion of workplaces with two interviewer sets would be about 27 per cent. As we have seen, a very large workplace may have as many as five interviewer sets. Thus the proportion of workplaces in the survey to which more than one interviewer is attached is almost certainly somewhere in the range 15-25 per cent.

The way in which employees are incorporated into interviewer sets appears to vary. On the one hand, it is sometimes stated that employees (including engineering-technical personnel) are surveyed at the same enterprises as workers, it being impractical to survey employees alone at an enterprise (Postnikov 1953, Ananyeva 1964). Why this should be impractical is unclear. An example of a mixed set of workers and employees

is given in Table B1.1 (note 4). On the other hand, although we have no evidence of pure employee sets at industrial enterprises, we do have evidence of such sets at non-industrial establishments. For example, Dedyulya (1972) refers to the coverage of employees at a construction administration in Minsk oblast' in Belorussia. Doctors, nurses ("intermediate medical personnel") and teachers are selected at medical establishments and schools situated near selected industrial enterprises and collective farms, a geographical restriction imposed "for organisational reasons" (Postnikov 1953, Krylov 1957). Pensioners do not constitute their own interviewer sets; they are worker and employee participants who are retained in the survey after their retirement (see Chapter B2).

Having explained the concepts of "interviewer set" and "workplace set", we may proceed to the method of selecting workplaces for the survey. In Table B1.2 we show the format of the lists from which enterprises are selected in a given branch and oblast'; in practice a list takes the form of a file of cards (kartochki), one card for each workplace (Posobie... 1980). In Table B1.3 we illustrate the selection method for a hypothetical branch and oblast', using a format adapted to facilitate explanation[3]. We are supposing that the engineering and metalworking branch in oblast' X consists of twenty enterprises and 40,000 production workers; staff occupied outside production are excluded from the survey (see Chapter B4). The distribution of enterprise sizes is concentrated around the average of 2,000 workers, but there are three very small enterprises and one very large enterprise. Were allocation of the sample in fact proportional, only about 40 workers (0.1 per cent) would have to be selected - that is, two interviewer sets of 20. However, the coverage of heavy industry is denser than average, and we shall assume that 125 workers are to be selected - five interviewer sets of 25.

Enterprises are listed in descending order of average wage of industrial-

TABLE B1.2

FORMAT OF LISTS OF INDUSTRIAL ENTERPRISES USED IN SELECTION OF SAMPLE

No.	Name of enterprise	Address of enterprise	Average wage of industrial-productive workers (rubles)	Average list count of industrial-productive workers (persons)	Cumulative total of industrial-productive workers (persons)	Number of workers to be selected (persons)	Wages fund for industrial-productive workers (rubles)
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)

Source: Ananyeva (1964 p.33)

TABLE B1.3

SELECTION OF ENTERPRISES IN THE "ENGINEERING AND METALWORKING" BRANCH IN OBLAST' X - HYPOTHETICAL ILLUSTRATION

No.	Enterprise	Average wage rubles/ month	No. of workers	Cumul.	Selection points	No. of budgets
(1)	(2)	(3)	(4)	(5)	(6)	(7)
1	"Red Hammer"	180	2500	2500		
2	"Sputnik"	160	1000	3500		
3	"Lenin"	160	9000	12500	4000, 12000	50
4	"Northern Metal"	160	2000	14500		
5	"Sverdlov"	150	2000	16500		
6	"Urals Steel"	150	2500	19000		
7	"Homeland"	150	1500	20500	20000	25
8	"Red Sickle"	140	2000	22500		
9	"New Dawn"	140	2000	24500		
10	"Smelter"	140	3000	27500		
11	"Victory"	140	1500	29000	28000	25
12	"Honour to Labour"	130	1500	30500		
13	"Fire"	130	2000	32500		
14	"Scrapworks No.1"	130	500	33000		
15	"Horizons"	130	1500	34500		
16	"Alloy"	120	2000	36500	36000	25
17	"Spark"	120	1000	37500		
18	"Forge"	120	1500	39000		
19	"Scrapworks No.2"	110	500	39500		
20	"Scrapworks No.3"	100	500	40000		

Number of selection points)) Number of interviewer sets)	5	Population mean wage	145.5
		Sample mean wage	146.0
Total number of budgets	125	Discrepancy	0.3%
Total population of workers	40000	Population standard deviation of wage	5.5
Sampling fraction	0.31%	Sample standard deviation of wage	5.3
Interval	8000		
Starting point	4000	Discrepancy	3.8%

productive workers (Table B1.2, column (4)), calculated by dividing the size of the workforce (column (5))[4] into the wages fund (column (8)). Cumulative totals of the number of workers in the top n enterprises on the list (n = 1, 2, 3, 4 ...) are calculated from column (5) and entered in column (6). Column (7) is blank at this stage of the exercise. Let us now turn to the adapted list in Table B1.3, where columns (3), (4), (5) and (7) correspond to columns (4), (5), (6) and (7) respectively of Table B1.2.

Systematic selection is used in the selection of enterprises (see Chapter A4). As the aim of the method is to achieve proportional representation of groups of workers with different wage levels, we do not select every k'th enterprise on the list (for some value of k) but those enterprises which contain every I'th worker on the implicit list of workers underlying the list of enterprises (for some value of the "interval of selection" I). The interval is calculated by dividing the number of interviewer sets (5) into the total number of workers in all the enterprises (40,000), giving 8,000. "Selection points" are determined on the implicit list of workers, starting in the middle of the first interval (4,000) and then adding successive multiples of the interval length (giving 12,000, 20,000, 28,000 and 36,000). The enterprises corresponding to the selection points are read off from column (5) and noted in column (6). One interviewer set is allocated to the enterprise containing each of the selection points. A very large enterprise, with a workforce exceeding the interval length, may contain more than one selection point, and thus is allocated more than one interviewer set. In Table B1.3, the "Lenin" plant (No.3 on the list), with 9,000 workers, is such an enterprise; it corresponds to two selection points at 4,000 and 12,000 and is therefore allocated two interviewer sets. We now see how a workplace set may contain two or more interviewer sets. Thus, 50 budgets are to be collected at the "Lenin" plant, and 25 each at

the "Homeland", "Victory" and "Alloy" plants (Nos. 7,11 and 16 on the list respectively).

As a check that the procedure has yielded a representative sample of workers as regards wage level, the average wage of the selected workers is calculated on the assumption that the average wage of the workers selected at any enterprise is equal to the average wage of all workers at that enterprise. This calculated average is then compared with the average wage of all the workers in the group of twenty enterprises. If the discrepancy between the two averages is less than three per cent (five per cent according to some sources), the results of the check are judged satisfactory. In Table B1.3 the discrepancy turns out to be only 0.3 per cent. If the discrepancy is not less than three (five) per cent a selected enterprise with average wage far from the population average is replaced by a new enterprise with average wage near the population average; the enterprise to be replaced and the new enterprise are chosen by trial and error so that the check can be satisfied[5].

In Table B1.3 we show another check of representativeness as regards wage which is rarely made in Soviet practice: a comparison of the standard deviation of the wage distribution in the population and in the sample. We see that the relative discrepancy here (3.8 per cent) is more substantial. The wage distribution of the sample is more concentrated than that of the population because the selection procedure prevents any representation in the sample of the extremes of the population wage distribution. We note that the method used to correct an excessive discrepancy between population and sample means greatly exacerbates the discrepancy between population and sample dispersions. We shall discuss this question more fully in Section 5.

We proceed to the selection of workers at the enterprises selected. Two separate lists of workers at each enterprise are compiled, one covering "qualified" workers and the other "little-qualified" workers. The

qualification category is determined by grade[6]. We shall see that a third category of the least skilled workers is excluded from the selection process altogether (Chapter B4). Each of the two lists is arranged in descending order of average monthly wage. The sample for the enterprise is divided into two subsamples, one for each qualification group, in proportion to the numbers of workers in the groups. The interval to be used in each list is calculated by dividing the number of workers to be selected from that list into the total number listed. Systematic selection is again used to select the workers whose family budgets are to be collected, starting from the middle of the first interval.

If the worker selected refuses to participate in the survey, a replacement is found, "paying attention to the maintenance of sample representativeness" (Posobie ... 1980). That is, the replacement should have wage, total family wage, occupation, qualification level, family size and composition as close as possible to those of the worker being replaced (Postnikov 1953). The same rule applies when replacing a worker who drops out of the survey for any reason. However, the rule is often not observed in practice (see Chapter B5).

3 Selection of families of collective farmers

Families of collective farmers, like families of workers and employees, are selected in two stages, with systematic selection used at each stage: first, selection of collective farms, and second, selection of households within each collective farm selected. Here, however, selection on the basis of workplace is equivalent to selection on the basis of place of residence, because collective farms are at the same time the workplaces and the areas of residence of the families which are their members.

The statistical office of each oblast' where collective farmers are surveyed compiles a list of all the collective farms in the oblast'. The

format of such a list is shown in Table B1.4. Collective farms of each "productive orientation" are grouped separately in the list. The groups are arranged in descending order of their importance for the agricultural economy of the oblast', with the "basic" group coming first (Matyukha 1966). There are three main productive orientations: crop-growing; industrial ("technical") crops such as cotton, flax and beet; and livestock-rearing. Less common orientations include "vegetable-potato", "garden-vineyard" and "citrus"[7]. Within each such group collective farms are arranged in descending order of amount of income distributed per labour-day or per person-day of work[8], as an indicator of the prosperity of the collective farm.

The initial selection of collective farms from the list follows lines analogous to those along which enterprises are selected. Selection points are located on the underlying list of households using the cumulative totals in column (5) of Table B1.4 and an interval length calculated by dividing the number of interviewer sets to be selected into the total number of collective farmer households in the oblast'. Systematic selection is again started at the middle of the first interval. There is no indication that more than one interviewer set is ever selected at the same collective farm.

The initially selected farms may not be very evenly distributed over the territory of the oblast'. As climatic and soil conditions, which have great influence over the economic situation of collective farms, may vary appreciably from one part of the oblast' to another, it is considered important to guarantee a fairly even territorial distribution. Another consideration here is that the opportunities open to collective farmer households for purchase of consumer goods and for use of collective farm markets (whether as sellers or as purchasers) depend on the distance separating their collective farm from nearby towns. If two or more collective farms in the same raion are initially selected, only one of them enters the

TABLE B1.4

FORMAT OF LISTS OF COLLECTIVE FARMS USED IN SELECTION OF SAMPLE

No.	Name of collective farm	Average money value of a work-day (rubles)	Number of households in the collective farm	Cumulative total of number of households	Name of raion	Note of whether the collective farm has been selected
(1)	(2)	(3)	(4)	(5)	(6)	(7)

Source: Ananyeva (1964 p.36)

final sample, and the others are replaced by collective farms adjacent on the list and situated in other raiony (Postnikov 1953, Krylov 1957; see also Chapter B3). For this reason the name of the raion in which each farm is situated is entered on the list (column (6), Table B1.4).

The final sample of collective farms should therefore be fairly representative of the population as regards territorial distribution as well as distribution by productive orientation and by per-capita net income. Nevertheless, the method does not always achieve its goal: Kildishev et al (1980 p.388) note that "unevenness in geographical distribution and in meteorological conditions may harm the representativeness of the collective farmer sample"[9].

In the selection of households at each of the collective farms selected, it is considered necessary to ensure that the sample is representative with respect to both of the two main components of household income: payment for work on the collective farm and income from the private plot. The number of labour-days or person-days worked by members of the household in the preceding year is considered a satisfactory measure of the former, and whether or not the household possesses a cow (or cows) a reasonable indicator for the latter (Posobie... 1980). Two separate lists of households are therefore compiled - one of households possessing at least one cow, and one of households without any cows (Kozlov et al 1965, Dolgushevskii and Khristich 1976). However, an unknown proportion of households do not fall into either of these categories, as they own a cow (or sometimes more than one cow) jointly with a neighbouring household, and these households are excluded from the survey (Shvyrkov 1965).

Each of the two lists is arranged in descending order of number of labour-days or person-days worked by family members in the preceding year, and systematic selection of households is carried out in the familiar fashion. The sample for the collective farm is allocated between the two

lists proportionally to the number of households in each, and the interval is calculated by dividing the number of households to be selected from the list into the total number of households on the list. Selection once again starts from the middle of the first interval.

Initially selected households which have no able-bodied members working on the collective farm are excluded from the sample and replaced by other households adjacent on the lists (Grankov 1955).

The checks made of the representativeness of the collective farmer sample are considered in Section 3 of Chapter B6.

If a selected family refuses to participate in the survey, or later drops out of the survey for any reason, it is replaced by another family similar to it with regard to size of family, type of work done by members on the collective farm, number of labour-days worked per year by members, number of privately owned livestock and size of private plot (Postnikov 1953, Krylov 1957). A family which is absent from the collective farm for more than two months is also replaced. If a family splits into two, of which only one remains on the collective farm, the remaining family continues in the survey; if both remain, only the one closest in its characteristics to the original family is retained. If a family in the survey combines with another not in the survey, the combined family continues in the survey. If two families, both in the survey, combine, they are both replaced by two new families, similar to the original ones (Dolgushevskii and Khristich 1976). These rules seem designed to minimise any sudden changes in the characteristics of participating families.

The data used to compile the lists of households are taken from the "economic book of village Soviet record-keeping" (pokhozyaistvennaya kniga selsovetskogo ucheta) (Grankov 1955, Annyeva 1964). This is a record of the basic demographic and economic characteristics of all the households in

a village (Dmitrieva 1980), but the quality of the records kept is very low (Shubina 1973, Sovershenstvovanie... 1979 pp.149-51). This must have deleterious effects on the sample drawn on the basis of these records.

4 Preliminary assessment of the methods of sample selection

Soviet writers routinely define the sample design used in the family budget survey as "stratified proportional sampling with systematic selection within strata" (Grankov 1955, Ananyeva 1964). We have already pointed out that the claim of proportionality is a false one. The formula cited also neglects to mention key aspects of the sample design: that coverage is incomplete, that stratification is primarily on the branch principle, that systematic selection starts from the middle of the first interval, and that the design incorporates clustering of the final sampling units in workplace clusters. The method described in the last two sections provides for the selection within each stratum of clusters (workplaces) in such a way that the probability of inclusion of any cluster in the sample is proportional to its size (as measured by the number of final sampling units - workers, employees or collective farm households - it contains). This is a recognised technique known as "probability-proportional-to-size (pps) sampling". We would therefore define the sample design as "incomplete non-proportional stratified sampling, with strata formed primarily on the branch principle, with systematic pps selection of clusters (workplaces) within strata and systematic selection of final sampling units (persons or households) within clusters, systematic selection starting from the middle of the first interval".

The use of cluster sampling is quite defensible. The high correlation among sampling units in the same cluster (intracluster correlation) reduces the efficiency of the sample in terms of precision achieved per sampling unit, but not necessarily its efficiency in terms of precision per unit cost,

because clustering makes data collection cheaper and more convenient. However, the correct estimation of sampling error requires that the fact that clustering is being used be recognised, and failure to recognise this leads to a considerable under-estimation of sampling error (see Chapter B6 Section 5).

Karapetyan, Rimashevskaya and Sidlyarenko (1967) comment that "the main sampling scheme used (in the FBS) is based on sampling theory", but that "the specific way this scheme is put into effect" is marked by serious deficiencies which vitiate the potential soundness of the basic scheme. These deficiencies suggest that those responsible for the organisation of the FBS have a poor understanding of sampling theory.

The deficiencies in the sample design of the FBS are listed in Table B1.5. Let us summarise them at this point.

1. Large parts of Soviet territory - for example, almost all of Eastern Siberia and the mountainous parts of Tadzhikistan - are not covered by the survey (Chapter B3), nor are important branches of the national economy such as food industry and trade (Chapter B4). We have seen also that other large groups - workers in industry outside production, unskilled workers, collective farmers who share cows, the disabled - are excluded.

Although a budget survey with incomplete coverage may have value nevertheless for purposes of academic analysis, it is severely deficient for the purpose of studying the Soviet population as a whole, for planning or policy-making. One cannot but be astonished at the defence of incomplete coverage by Matyukha (1967):

> Because not all population groups are usually covered by the budget survey, (its) data inadequately represent corresponding indicators of the national population of workers, collective farmers and employees. However, is this a shortcoming of contemporary budget surveys? No. It is in general incorrect to set such a task before budget surveys in the conditions of our country.

TABLE B1.5

DEFICIENCIES IN THE SAMPLE DESIGN OF THE BUDGET SURVEY

Necessary conditions for the design to be sound	Actual situation	Where in thesis discussed
1. The whole population of the USSR should be covered.	Many important gaps in coverage - by social group - by territory - by branch.	 Chapter B2 Chapter B3 Chapter B4
2. Either: representation of different population strata should be proportional,	Principle of proportionality affirmed but far from being implemented.	Chapters B2, B3, B4
or: data should be adequately reweighted to compensate for disproportionalities.	Situation complex and unclear; reweighting probably exists but is inadequate.	Chapter B9
3. Stratification variables should be substantively meaningful*.	Key stratification variables meaningless	Chapters B10, C3
4. Sample should undergo steady renewal	Sample very rarely renewed	Chapter B1 Section 7, Chapter B5
5. Either: the sampling unit and the observation unit should be identical.	Sampling unit is working person, observation unit is family.	Chapter B1 Section 6
or: special provision must be made so that no observation unit can enter sample in more than one stratum.	No such provision made	
6. In systematic sampling, the ordering variables should be substantively meaningful*.	Key ordering variables not very meaningful.	Chapter B1 Section 4
7. In systematic sampling, start should be random (or change systematically)	Fixed start at middle of first interval	Chapter B1 Section 5
8. Checks of sample representativeness should be adequate.	Checks inadequate	Chapter B6
9. Most important test of soundness of design: income distribution of sample should be close to that of the population	Badly fails test.	Chapter B7

2. There is no necessity for the allocation of a stratified sample among its strata to be proportional. Some of the disproportionalities in the survey design have justifiable rationales, though many others do not. Thus the disproportional allocation of the collective farmer sample among oblasti is to a large extent motivated by a desire to ensure minimum sample sizes in even the least populated oblasti in order to provide a reliable basis for calculations at oblast' level (see Chapter B3). However, care must be taken to reweight data to compensate for all disproportionalities. We are assured that weights are applied to oblast' data on collective farmers, but the situation as regards the reweighting of data on workers and employees with corrective coefficients appears to be unsatisfactory (see Chapter B9).

3. As one purpose of stratification is to ensure that data are representative within each stratum as well as globally, it is desirable that strata be defined in a substantively meaningful way. A recurring complaint of critics is that key stratification criteria used in the survey - "social group" (workers, employees, collective farmers) and branch of the economy - classify families in an arbitrary way, yielding data in breakdowns of no real interest to analysts (see Chapter C4). The problem arises mainly from the organisation of the survey on the branch principle (Chapter B10).

Irrelevant stratification criteria, though undesirable, do not in themselves prevent the sample from being globally representative, provided that the criteria unambiguously allocate each surveyed unit to one and only one stratum. It is this condition which the budget survey fails to fulfil (point 5).

4. The long periods over which families participate in the survey, and the absence of provision for adequate renewal of the sample, entail the

biases discussed in Section 7 of this chapter.

5. In the survey of workers and employees, the final sampling unit of the selection procedure is the individual worker or employee, while the surveyed unit is the family of which this individual is a member. As a result of this discrepancy between the sampling unit and the surveyed unit, any given surveyed unit (family) can enter the survey through the selection of any one of the sampling units (working family members employed in branches covered by the survey) it contains. Usually family members work in different branches, and the classification of the family under one branch rather than another is then a matter of arbitrary chance. Furthermore, the greater the number of sampling units (zero, one, two, three or more) a given surveyed unit contains, the greater the probability of its inclusion in the sample. This is the "multi-worker bias" discussed in Section 6 of this chapter.

Soviet critics often argue that this bias must be overcome by making the sampling unit identical with the surveyed unit, elevating the identity of the two to a fundamental principle of sampling theory. "Selection should be not by working members of families but by families. The sampling unit and the surveyed unit must become the family" (Shvyrkov 1965). This is to be achieved by reorganising the survey on the territorial principle. Although this reorganisation is desirable on several grounds, it would be quite possible to eliminate the multi-worker bias while retaining the branch principle[10]. All that is required is to make sure that any given surveyed unit can enter the survey through the selection of one and only one sampling unit. Thus one could delete from the lists of workers and employees all those who were not the largest contributors to the income of their families. But no provision of this kind is made in the design of the Soviet survey.

6. The purpose of the systematic sampling which forms part of the survey design is to provide for the sample to have a representative distribution with respect to the variables used to order the lists. It is therefore desirable that these variables be such that the attainment of a representative distribution with respect to them be a worthwhile goal, either because they are substantively important variables or because, if not important in themselves, they are highly correlated with more important variables of which they may be regarded as proxies.

Official apologists of the survey claim that the wage of the individual worker or employee, used as the ordering variable, is just such a proxy for family income, the really significant variable (Matyukha 1967, Posobie... 1980). This, they argue, follows from the fact that "wages are the main source of (family) income". As the critics object, the point is specious. Total family wage is of course the main source of family income, but the ordering variable is not total family wage but the wage of an individual family member often accounting for only a small proportion of the family income. It may be, for example, the wage of a nurse daughter living with parents earning high wages. Besides, the standard of living of a family is determined not so much by total family income as by per-capita income, reflecting the ratio of working members to dependents (Karapetyan, Rimashevskaya and Sidlyarenko 1967)[11].

Similar criticisms are made of the ordering variables used in the survey of collective farmers. Number of labour-days worked and possession or non-possession of a cow, even taken together, are "far from a complete indicator of material welfare" (Shvyrkov 1965).

7. Systematic selection in the FBS sample design starts from the midpoint of the first interval, excluding extreme units from the possibility of selection. The bias caused by this practice has already been discussed

in Chapter A4. Its consequences for the FBS sample are examined in Section 5 of this chapter.

8. TsSU checks the representativeness of budget survey samples by a variety of methods, considered in Chapter B6. Critics complain that the checks are made with respect to too restricted a range of variables, some of which lack substantive significance. For example, the criticisms made of wage of the selected individual as an ordering variable apply equally to it as a check variable. Representativeness checks focus almost exclusively on means, so that the representativeness of other measures of the distribution of sample variables (above all, that of dispersions) is neglected, and measures taken to improve representativeness with respect to means exacerbate bias with respect to dispersion. The problem is related to the little use made of sampling theory in making the checks.

9. The most crucial requirement of a sample in a budget survey is that it be representative with respect to the distribution of families by per-capita income. The income distribution of the budget survey sample is very badly biased: mean income is subject to upward bias while dispersion of incomes is grossly underestimated through the over-representation of middle-income families in the sample and the under-representation of both low-income families and families with very high incomes. These biases are the overall net effect of biases arising from many individual deficiencies in sample design (see Chapter B7). They provide a measure of the poor quality of the sample design taken as a whole.

5 "Tail-cutting bias"

In this Section we examine the biases which arise from systematic sampling with a mid-interval instead of a random start. However, we should

first note that the use of a mid-interval start in fact represents an advance on former practice. Until 1958 systematic sampling in the FBS started not from the midpoint but from the beginning of the first interval (Ananyeva 1964)[12]. This caused a significant downward bias in sample means.

Of all possible fixed-start rules, the mid-interval start minimises bias in the sample mean. It does this by ensuring the equality of population and sample medians: that is,the same point divides the population and the sample wage distributions into two parts containing equal numbers of workers (to take one case from the FBS). In a symmetric distribution mean and median coincide. If the wage distribution is not too asymmetric, therefore, mid-interval starting does not excessively bias the sample mean. As in Table B1.3, the representativeness check for the mean is satisfied. But significant bias can arise if the population distribution of the ordering variable is markedly asymmetric (Venetskii and Chernysheva 1978)[13].

The problem of bias in the sample mean turns then on the degree of asymmetry of wage distributions. In the USSR as in the West, these tend to take a lognormal form, which is of course asymmetrical (Rabkina and Rimashevskaya 1972). Budget survey samples, however, are not drawn from a population which spans the full range of wages existing in the society, but from one part only of the population which includes neither those on the lowest wages nor those on the very highest. Such a truncated distribution will be much less asymmetrical, and we suspect that the bias in the sample mean due to asymmetry is not a very serious one. It could become more serious if the income distribution of the sample were made more representative with regard to dispersion, while the method of systematic sampling remained unchanged.

In Table B1.6 we explore the consequences of mid-interval systematic

selection by drawing samples of different sizes, from one enterprise up to six, from the hypothetical list of twenty enterprises in Table B1.3. For purposes of illustration, we again make the simplifying assumption that the workers selected at any enterprise all have the average wage for that enterprise. We note that the bias in the sample mean may often turn out much larger than in Table B1.3, though none of the variants would clearly fail a standard Soviet representativeness check[14]. We suggest that the pattern of biases in the sample standard deviation can be largely explained in terms of the combined action of two tendencies:

(a) Other things being equal, the smaller the sample, the more of the tails at the two extremes of the population distribution is excluded from the sample, and thus the smaller the standard deviation of the sample distribution: the "tail-cutting bias".

(b) Samples containing an even number of enterprises contain no enterprise close to the centre of the population distribution, increasing the sample standard deviation, while the reverse is true of samples containing an odd number of enterprises. This effect is the stronger the smaller the sample, and we see that in the two-enterprise case it more than compensates for the effect of (a).

In three out of five cases, the relative discrepancy between population and sample standard deviations would exceed the upper tolerance limit (5 per cent) cited for checks of sample representativeness, were such checks applied to standard deviations. This demonstrates the unreliability of the selection method used for preserving the dispersion of the population distribution in the sample, at least for small samples. Both tendencies (a) and (b) would be attenuated in sufficiently large samples, though much depends on the exact shape of the population distribution - for example, the length of the upper tail.

It appears that samples of enterprises selected from <u>oblast'</u> -branch

TABLE B1.6

CHARACTERISTICS OF SAMPLES OF DIFFERENT SIZES FROM THE HYPOTHETICAL POPULATION OF ENTERPRISES USED IN TABLE B1.3

No. of selection points		1	2*	3	4	5**	6
No. of budgets	min.	20	40	60	80	100	126
	max.	25	50	75	99	125	150
Sampling fraction per '000	min.	0.5	1.0	1.5	2.0	2.5	3.1
	max.	0.6	1.25	1.9	2.5	3.1	3.75
Interval length		40000	20000	13333	10000	8000	6667
Selection points		20000	10000	6667	5000	4000	3333
			30000	20000	15000	12000	10000
				33333	25000	20000	16667
					35000	28000	23333
						36000	30000
							36667
Numbers of the enterprises selected		7	3	3	3	3	2
			12	7	5	3	3
				15	10	7	6
					16	11	9
						16	12
							17
Population mean wage		145.5	145.5	145.5	145.5	145.5	145.5
Sample mean wage		150.0	145.0	146.7	142.5	146.0	143.3
Discrepancy		+3.1%	-0.3%	+0.8%	-2.1%	+0.3%	-1.5%
Population standard deviation of wage			5.5	5.5	5.5	5.5	5.5
Sample standard deviation of wage			6.7	4.8	5.4	5.3	5.2
Discrepancy			+22.0%	-12.2%	-1.8%	-3.8%	-5.5%

* This sample size corresponds to the average sampling fraction of the budget survey: 1 worker in every 1000.

** This is the sample size illustrated in Table B1.1.

lists usually consist of few enterprises, and that the selection of a single enterprise from a list is not rare. For example, according to Vladykin (1955), ten enterprises are surveyed in Kuybyshev oblast' in four branches: three engineering plants, two chemical plants and five enterprises in the two other branches. In Tadzhikistan we have just one enterprise surveyed in the metalworking branch, one clothing factory and three textile factories (NKh TadzhSSR v 1961g). Thus samples of enterprises would usually not be large enough greatly to attenuate the effects we have been discussing. The enterprises in any branch and oblast' paying the best and the worst wages, for example, would almost always be excluded from the sample. Lists of collective farms, on the other hand, tend to be much longer, minimising this kind of bias.

The sample of workers selected within an enterprise (or of collective farmer households within a collective farm) is probably large enough to minimise tail-cutting bias. At the bottom end of the income distribution, tail-cutting bias is overridden by more substantial biases such as those arising from the exclusion of unskilled workers and collective farmer households without able-bodied members. We might note, however, that the tail-cutting bias excludes the possibility of the people with the highest salaries in any enterprise or establishment - the Director, the Chief Engineer and so on - being included in any sample of employees.

6 "Multi-worker bias"

We have already explained the nature of the "multi-worker bias". Volkov (1962) gives an algebraic analysis of it based on the simplifying assumption that the sampling fraction used for the selection of workers and employees in all branches is p. A family enters the sample if any one of its working members is selected, and the selection of any member is assumed independent of the selection of any other member. The probability

of a family with just one worker entering the sample is simply p. If a family has $n \geqslant 2$ workers, the probability of no member being selected is q^n, where $q = (1 - p)$. Therefore the probability of the n-worker family entering the sample is $(1 - q^n)$. Volkov presents illustrative calculations for the incomes survey of 1958 (Table B1.7), which was organised on similar lines to the budget survey but covered all non-agricultural branches of the economy. We see that the probability of a family entering this survey is approximately proportional to the number of working members. In the budget survey the relationship is complicated by the variation of the sampling fraction among branches and territories.

The standard Soviet family with two working parents and children (if any) below working age has a probability of inclusion in the survey intermediate between the lesser probability of inclusion of (say) a single-parent family with one worker and the greater probability of inclusion of a family with three or more workers. Families with three or more workers tend to consist of working parents, not yet of retirement age, living with their working children and children-in-law. They are characterised by relatively high total family income and by a high ratio of workers to dependents. Thus the multi-worker bias entails considerable biases in the distribution of the sample by family or per-capita income and by types of sex-age family composition.

Soviet writers propose three possible methods of eliminating multi-worker bias (Volkov 1962, Karapetyan 1970). It could be compensated for by reweighting the data; provision could be made so that each family could enter the sample through only one of its members; or the survey could be reconstructed on a territorial basis.

From his algebraic analysis Volkov concludes that the appropriate factor for reweighting the data for a family with n workers is $p/(1 - q^n)$, which, p being very small, can be approximated by $1/n$. As the analysis i

TABLE B1.7

UNEQUAL PROBABILITIES OF INCLUSION IN THE INCOMES SURVEY OF FAMILIES WITH DIFFERENT NUMBERS OF WORKERS

Probability of inclusion in the incomes survey of September 1958 of a family with one worker = p = 0.005

$q = 1-p$ = 0.995

Probability of inclusion of a family with n workers = $1 - q^n$

Ratio of probability of inclusion of a family with n workers to probability of inclusion of a family with one worker

= $(1 - q^n)/p = R \simeq n$ for small p.

n	q^n	$1-q^n$	R
1	0.995	0.005	1
2	0.990025	0.009975	1.995
3	0.985075	0.014925	2.985
4	0.980150	0.019851	3.970

Source: Volkov (1962)

TABLE B1.8

TRIAL CALCULATIONS WITH THE "HEADS OF FAMILIES" METHOD ON INCOMES SURVEY DATA FOR ARMENIA IN 1958

	With a sample of heads of families only (base = 100)	With the full sample actually used	
		before correction	after correction
Average number of workers per family	100	116	100
Average number of persons per family	100	106	99
Average total income of family	100	113	102
Average per-capita income of family	100	109	104

Source: Karapetyan (1970)

based on the assumption of uniform sampling fraction in the selection of workers and employees, this procedure would preserve the biases due to disproportional coverage of branches and territories. It would also be very inconvenient to organise.

The second method requires that data be collected on the family composition of workers and employees on the selection lists. The sample would be selected systematically as at present, but any who were not heads of families[15] would be excluded from the sample (Karapetyan, Rimashevskaya and Sidlyarenko 1967). In fact it would suffice to collect data on family composition from the group initially selected, though even this of course would involve additional expense and effort. This sort of procedure is used in the budget surveys of Poland (Shvyrkov 1965), of the GDR (Kazun 1976) and of Rumania (Ananyeva 1966). It "is not a radical solution to the problem, but merely a way out of artificially created difficulties" (Shvyrkov 1965).

Calculations by Karapetyan (1970) on the multi-worker bias and methods of dealing with it, using data relating to Armenia from the 1958 incomes survey, are of interest (Table B1.8). The results of the "heads-of-family" method, which fully eliminates the bias, serve as a standard of comparison. We see that the multi-worker bias entailed for these data an upwards bias in family income of 13 per cent. Correction by reweighting did not fully eliminate the bias.

Rimashevskaya (1965) reports other calculations on the multi-worker bias, probably conducted on the same data by the Scientific Research Institute of Labour (NII-truda). The survey forms were reprocessed to exclude data relating to workers who were not heads of families. Comparison of distributions by family size, with and without reprocessing, with the distribution given by the population census confirms that the "heads of family" method does greatly reduce the bias (Table B1.9). Lacking any information on the sample upon which the survey distributions are based,

TABLE B1.9

EFFECT OF MULTI-WORKER BIAS AND OF THE "HEADS OF FAMILY" METHOD FOR CORRECTING IT ON THE DISTRIBUTION OF 1958 INCOMES SURVEY DATA BY FAMILY SIZE

No. of persons in family	Distributions by family size		
	Income survey data		Population census data
	including non-heads of family	excluding non-heads of family	
(1)	(2)	(3)	(4)
	%	%	%
1	13	20	20
2	18	21	21
3	26	25	22
4	22	19	18
5	12	9	10
6 or more	9	6	8
	100	100	100

Source: Adapted from Rimashevskaya (1965)

Method: The distributions given by Rimashevskaya exclude one-person families, but she also gives the proportions of single-person households in each of the three data sets, so that we are able to retrieve the complete distributions.

The time and place to which these incomes survey data refer are not specified. Rimashevskaya does, however, identify other distributions shown in the same source as taken from the 1958 survey, and she has also revealed that they relate to Armenia alone (Shenfield 1982b). These are then probably also 1958 data for Armenia.

we cannot judge whether the remaining discrepancies with census data are attributable to sampling error.

No doubt the territorial reconstruction of the survey, discussed in Chapter B10, would be the most advantageous method of doing away with the multi-worker bias.

7 Biases due to excessive periods of participation

Households take part in the British Family Expenditure Survey for a fixed period of one year, one quarter of the sample being replaced every three months on a rotating basis (Kemsley 1968). Such limited periods of participation are typical of Western surveys, and even shorter periods are common - only one week in the French survey. By contrast, as we have seen, large-scale sample formation is a rare occurrence in the Soviet budget survey: the most recent occasion was in 1969, and before that in 1951-2. Even these exercises involved mainly expansion rather than replacement of the existing sample. Efforts are made to "preserve" the sample - that is, to retain the same families in the survey for as long as possible (see Chapter B5). Families who nevertheless do drop out of the survey for one reason or another are replaced on a piecemeal basis.

It appears to be regarded as a desirable ideal for families to continue in the survey until the selected working member retires. Thus, at a meeting of survey participants in Moscow, the participant L.G.Starostina proclaimed:

> We are glad to help our State by our work, and call upon all those present never to stop keeping budget records and to keep them from year to year (Panina 1974).

"Veterans" of the budget survey - as the Deputy Head of TsSU Kirgizia calls long-term participants (Altunina 1974) - receive special accolades and honours: reports of meetings of participants preface speeches with a note of how many years the speaker has been in the survey (e.g. Soveshchanie...

1964, Opyt... 1955).

Two of these reports give statistics of length of participation. Of 1,800 families surveyed in Moscow in 1963, 330 (18 per cent) had been in the survey since 1952 and 25 (1 per cent) since the 1930's or 1940's (Sobranie... 1963). Of 830 families represented at a meeting in Lvov oblast' in 1972, more than 200 (24 per cent) had been in the survey at least since 1952 (Babayev 1972). Thus there is a substantial core of veterans comprising perhaps 15-25 per cent of the sample.

TsSU instructions specify that a family leaves the survey on moving to another population point or on a change of workplace leading to no family member working at an enterprise covered by the survey (Instruktsiya... 1960). Rimashevskaya states, on the basis of recalled conversation with Matyukha, that these are the main reasons people leave the survey, with about one-third of the sample having to be replaced every five years (Shenfield 1982b). This implies a median period of participation of nine or ten years. Thus there must be many families at the opposite end of the continuum to the veterans, who leave the survey after "only" a few years of participation. Retirement, illness and death are other recognised reasons for leaving the survey. A less readily admitted phenomenon, the importance of which we assess in Chapter B5, is refusal by families to remain in the survey.

One consequence of long-term participation in the survey is thought to be that families become atypically organised and regular in their budgetary behaviour. We consider this "guinea-pig effect" in Chapter B5.

Another consequence is the gradual divergence in the composition of the sample away from that of the population over time, "the drift of a stationary sub-population away from the population as a whole" (Karapetyan 1980). Here we put on one side the biases built into the sample at the time of its formation, which the routine representativeness checks conducted

by TsSU do not show up (Chapter B6), in order to focus on the new biases which arise later. Thus the checks made on the families selected at the end of 1951 gave satisfactory results (Postnikov 1952), but the same checks were showing substantial biases by the 1960's (Kozlov et al 1965).

The structure of the population changes over time in many significant ways which are inadequately reflected in the structure of the budget survey sample:

i. change in the breakdown of the population by "social group". In recent times the proportion of workers and employees, and the proportion of pensioners, have risen, while the proportion of collective farmers has fallen (Korovkin 1969).

ii. change in the branch structure of the economy. The proportion of workers and employees employed in the "non-productive sphere", under-represented in the survey, has increased. The development of new industrial regions entails change both in the branch structure of industry and in its territorial distribution (Korovkin 1969), but there is a very long delay before these regions are included in the survey. For example, the new industrial zone along the Baikal-Amur Railway (BAM) is not covered (Shenfield 1982b).

iii. change in the institutional structure of agriculture. Matyukha notes that the amalgamation of smaller into larger collective farms, and the conversion of collective farms into State farms, are inadequately reflected in the sample (Krasnoshchekov 1962; see also Chapter B4)[16].

iv. Parfenova comments that the collective farmer sample becomes less representative as the level of economic development of different collective farms changes over time (Sukhoruchkina 1970).

These are some of the ways in which the structure of the population changes while that of the sample lags behind. There are other respects in which the structure of the population does not drastically change but that

of the sample does, just because there is little turnover in the families comprising it. The sample literally ages: the average age of the working members of the families in the survey rises well above the population average. The lack of renewal of the sample results in serious under-representation of newly married couples:

> In the population new families separate from old ones, but in the sample the old families remain while the new young families are not surveyed (Kildishev et al 1980).

As moving to another enterprise usually involves leaving the survey, long-term survey participants have accumulated atypically long periods of service (stazh) at the same enterprise. As qualification level tends to rise with length of service, and as long-service workers receive much more in bonuses than short-service workers, this is associated with the earning of incomes well above the enterprise average (speech by Gavrilov, in Sukhoruchkina 1970). Because of high labour turnover, the under-represented category of workers who have been at an enterprise for only a short time constitute a large proportion of the workforce. Older people are also more likely to have children and children-in-law of working age living with them, so that the age bias exacerbates the multi-worker bias and the family wage is even more atypically high than the individual wage.

As sample turnover is nevertheless fairly substantial, the extent of bias greatly depends on the method used to replace families who leave the survey. The standard TsSU instructions require that the replacement worker or employee be chosen to have the same occupation, skill level, approximate wage, total family wage and total family income, and if possible family structure, as the person being replaced. (Average incomes over the past three months are used.) The proposed replacement is recorded on a special Form Z, showing the basic indicators of both old and new families, which has to be submitted to the head of the statistical administration for confirmation (Instruktsiya... 1960). There are some complaints, however, that the

instructions on replacements are not fully observed:

> Unfortunately instructions are often not followed, leading to worsened sample representativeness in a number of Republics (Matyukha, in Soveshchanie... 1960; see also Postnikov 1952).

Even when observed, these instructions merely conserve sample biases which have accumulated in the indicators taken into account. It is with respect to indicators, such as age, omitted from Form Z that sample turnover may serve to reduce bias somewhat. In fact the practice has been widely adopted in recent years of "directed selection" of replacements, guided by the aim of improving sample representativeness (Chapter B8). According to Rimashevskaya, this has helped achieve a considerable recent improvement in sample representativeness (Shenfield 1982b).

She also points out that some changes in the population over time tend to reduce the biases in the sample. In particular, the differential in average wages between the heavy-industrial branches well represented in the sample and the light-industrial and "non-productive" branches poorly represented or not represented in the sample has been narrowing. In order to verify this point, we have calculated the ratios of average wages between selected branches of heavy industry and selected branches of light industry at different times, as well as between industry as a whole, relatively well represented in the survey, and branches of the economy less well represented: trade and residential services (both excluded from coverage altogether) and agriculture (workers and employees on State farms). We have also compared rail transport, covered by the survey, with transport as a whole, not covered except for rail. The results, shown in Table B1.10, with two exceptions (chemicals: textiles and industry: residential services) confirm Rimashevskaya's point. They also show that differentials remain substantial, and so it is likely that biases remain substantial too.

One solution often suggested to the problem of long periods of

TABLE B1.10

RATIOS BETWEEN AVERAGE WAGES IN BRANCHES OF THE ECONOMY WELL COVERED IN THE FAMILY BUDGET SURVEY AND AVERAGE WAGES IN BRANCHES POORLY COVERED OR NOT COVERED

	Engineering and metalworking: clothing workers (1)	Ferrous metallurgy: food industry workers (2)	Chemicals: textiles workers (3)
1950	1.72	1.87	1.28
1955	1.66	1.70	1.27
1960	1.59	1.67	1.31
1963	1.44	1.50	1.26
1966	1.36	1.44	1.24

	Electrical and thermal energy: footwear workers (4)	Coal mining: forestry, wood and paper workers (5)	All industry: agriculture[1] workers (6)	workers and employees
1950	1.56	2.01	1.91	1.84
1955	1.35	1.69	1.71	1.68
1960	1.25	1.92	1.72	1.69
1963	1.29	1.85	1.48	1.47
1966	1.20	1.82	1.35	1.34
1980				1.25

(1) workers and employees of State farms and subsidiary agricultural enterprises

/cont'd...

TABLE B1.10 (cont'd)

	All industry: trade[2] workers and employees (7)	All industry: residential services workers and employees (8)	Rail transport: all transport workers and employees (9)
1950	1.50	1.43	1.03
1955	1.50	1.50	1.01
1960	1.56	1.59	0.95
1963	1.53	1.57	0.95
1966	1.35	1.41	0.92
1980	1.34	1.39	0.94

(2) including also public dining, procurements and material-technical supply

Sources: Figures for average wages in 1950-66 were taken from Trud v SSSR (M,1968) pp.138-45, and figures for 1980 from USSR in figures for 1980 (M,1981) pp.164-5. Ratios were calculated by us.

The handbook for 1980 gives data only by branch of the national economy (industry, construction, trade etc.) and not by branch of industry.

participation is a complete reconstruction of the sample at regular intervals, at least once every ten years (Shvyrkov 1965, Karapetyan, Rimashevskaya and Sidlyarenko 1967, Safronova 1968). The more radical solution would be to use very much shorter standardised periods of participation, as in the "method of momentary observations" (Chapter B11).

8 Conclusion

The FBS uses a stratified sample design organised on the branch principle, with workplace clustering and an imperfect form of systematic selection within strata and clusters. Although the basic framework of the design is informed by sampling theory, the design is marred by serious deficiencies. Apart from the biases which arise from the incomplete and uneven coverage of the population, three types of bias are inherent in the organisation of the sample: (a) biases caused by the erroneous application of systematic sampling, in particular the "tail-cutting bias"; (b) the "multi-worker bias" associated with the discrepancy between the sampling unit (the individual worker or employee) and the unit of observation (the family); and (c) various biases consequent upon the excessive periods of participation of families in the survey.

Notes to Chapter B1

1 Whether the FBS covers single-person households as well as multi-person families is seldom made clear. It appears that they are included in the survey of workers and employees but not in the survey of collective farmers (Kildishev et al 1980 p.383). Rimashevskaya (1968b) states that single-person households are under-represented in the FBS.

2 This seems the most likely procedure; sources differ on the details. Thus, Matyukha (1967 pp.11-12) states that the global worker sample is first allocated by branch for the USSR as a whole and then by territorial units, but Posobie... (1980) gives the two steps in the reverse order. The Deputy Head of TsSU USSR OBS, V.A.Samoilov, states that it was the Republican TsSU's which selected workplaces and workers for the additional sample at the end of 1968 (Sukhoruchkina 1970), but Parfenova refers to this work being done by oblast' offices (Soveshchanie... 1969).

3 We model our illustration on those used in such Soviet texts as Krylov (1957) and Matyukha (1967), but use artificially rounded figures to make the calculations easier to follow. Also, the calculations of standard deviation at bottom right of Table B1.1 do not appear in Soviet textbooks.

4 The average list (spisochnyi) count (Table B1.2, column (5)) is the average number of workers on the list of permanently established enterprise staff over the preceding year.

5 The checks of sample representativeness which are incorporated into the process of sample selection are fully discussed in Section 2 of Chapter B6.

6 This is the account given by all sources except Grankov (1955), who states that qualification category is determined either by grade or by occupation.

7 Before 1951 the survey took no account of the productive orientation of collective farms (Postnikov 1951).

The productive orientation of collective farms is identified from their annual reports.

Productive orientations which are uncommon in the USSR as a whole may be important in particular areas. For example, "natural and economic conditions in Georgia are very varied" (Morozov 1961), with thirteen agricultural zones and subzones of different productive orientation - grain-livestock, vineyard, subtropical crops, mountain-livestock etc.

Berzkaln (1968) appears to say that collective farms are sometimes grouped by economically relevant variables other than productive orientation - for example, by climatic or soil zones. As there is no other reference to such a practice, this should probably be interpreted as a recommendation rather than a description of actual practice.

8 Under the labour-day system for paying collective farmers, now superseded, residual net income of the collective farm after meeting all obligations was distributed among the members in proportion to the number of labour-days (trudodni) worked. One person-day of work could be equivalent to one or more labour-days, depending on the type of work.

9 Krylov (1957) gives us the remarkable information that the lists of selected collective farms are discussed and confirmed by raion Soviet executive committees.

Berzkaln (1968) reports work done by the Latvian Branch of NII TsSU at the Computer Centre of TsSU Latvia to develop algorithms for the computerised selection of samples of collective farms. The method is more thorough and formalised than, but not basically different from, that used in non-computerised selection. It is explained in Chapter B6. This is the only reference known to the use of computers in sample selection, as distinct from their use in data processing. There is no evidence for more than experimental use of the computerised procedure.

10 Similarly, critics often attribute the incomplete coverage of the survey to its organisation on the branch principle (Karapetyan, Rimashevskaya and Sidlyarenko 1967). But it would be quite possible to cover all population groups on the branch principle, with separate coverage of those outside employment (pensioners etc.).

11 Perhaps an even better indicator of the standard of living than per-capita family income would be per-yedok family income. The yedok was a consuming unit widely used in Soviet budget surveys of the 1920's as an "adult equivalent", with children of different ages counted as different fractions of a yedok. Postnikov (1961) has advocated the revived use of the unit.

12 Mid-interval starts had in fact been laid down in the instructions for the budget survey for 1950, but starting from the beginning of the interval was once again specified in the instructions for 1952 (Krylov 1957 pp.76-77). The improvement was secure only from 1959.

13 We are not arguing that asymmetry leads to bias in the mean, but only that it removes the guarantee of an unbiased mean provided by symmetry.

Matyukha (1967) states that "it not rarely happens that, as a result of disregard for the principles of proportional stratified sampling with systematic selection within strata, the selected enterprises turn out to be unrepresentative". Rules may be broken, but Matyukha, in his misplaced confidence in the soundness of the principles applied, probably is attributing discrepancies which are inevitable if the rules are followed to breaking of the rules.

14 That is, discrepancies would not exceed the tolerated level of 3-5 per cent. Our use of the standard criteria of representativeness employed in the checks is a matter of convenience, and does not imply that they make sense from a sampling theoretic point of view.

15 Alternatively, any who were not the "basic worker" of the family could be excluded. "Only force of custom can explain the survival of the outdated concept 'head of family'. The main person representing the family in statistical tabulations should be the person earning the greatest income" (Karapetyan 1980).

16 There seem to be difficulties even in adapting the organisation of the survey to changes in the boundaries of oblasti. "TsSU Ukraine has not yet studied the question of changing the collective farmer samples in Odessa and Lvov oblasti, in spite of significant changes in their boundaries" (speech by Matyukha, Krasnoshchekov 1962).

APPENDIX 1 TO CHAPTER B1

USE OF THE FBS SAMPLE FOR THE CONDUCT OF OTHER SURVEYS

The FBS is supplemented by a range of one-off surveys which are used to collect additional information on a variety of topics - household inventories of consumer durables, time budgets, everyday services, living conditions, the use of public dining establishments, the availability of pre-school childcare, fertility etc. (Ananyeva 1964, Karapetyan, Rimashevskaya and Sidlyarenko 1967, Posobie... 1980). Most of these surveys are conducted "on the budget network" - that is, using subsamples of the FBS sample. In that way they can most easily be fitted into the work programme of the TsSU Department of Budget Statistics.

Ananyeva (1964) lists four major supplementary surveys conducted in the few years preceding:

(a) a survey of the living conditions of workers and collective farmers;

(b) a survey of the wages of workers and employees by occupation;

(c) a survey of natality in the families of workers, employees and collective farmers; and

(d) the 1958 survey of family composition, incomes and living conditions.

We shall dwell a little on the last two of these; (c) is a typical example of a supplementary survey "on the budget network", while (d) was the first in an important series of surveys with special characteristics.

The survey of natality was conducted in 1960 "on the basic budget network" (Kozlov et al 1961 p.228). This means that only those families were interviewed who had been participating in the FBS continuously from the beginning of 1959 up to the time of the natality survey in 1960, that is, they were well established as participants. There were 37,000 such families, within which the women aged seventeen and over, numbering 54,500,

were interviewed. We have no explanation of why the survey was restricted to such families; perhaps families who had only recently entered the FBS were considered to be less accustomed to the burden of participation, and so in many cases less willing and reliable sources of information.

The questionnaire of the natality survey comprised three sections:

A. General information on the family (income, living conditions etc.);

B. Information about the woman interviewed (length of service at work, age at marriage, number of years married etc.); and

C. Information about children born to the woman interviewed.

The same interviewer responsible for collecting the family budget for the budget survey administered the questionnaire for the natality survey, filling in Section A from data already collected for the budget survey and Sections B and C by interview on one of the routine visits to the family required by the budget survey. Thus no special interviewing arrangements are necessary for such supplementary surveys; use is simply made of the existing infrastructure of the budget survey.

The data from the survey were used to study the relationships between natality on the one hand and the age of the woman, her employment status, family living space per person and per-capita family income on the other.

The "survey of family composition, incomes and living conditions of workers and employees in the non-agricultural branches of the national economy", first conducted in 1958, became a fairly regular exercise, carried out about once every three years. Following Czech practice with regard to the similar survey conducted in Czechoslovakia, we shall refer to this survey as the "microcensus" for the sake of brevity. The microcensus differs from other supplementary surveys in using not only the whole of the FBS sample (except, before 1972, the agricultural part of it) but a large supplementary sample as well, so that the total sample is about five times

as large as the FBS sample - 240,000 families in 1958, 280,000 in 1967 and 310,000 in 1972 and 1975 (McAuley 1977 p.219, Vsesoyuznoe... 1969 p.299, Vorontsova 1975). The TsSU Departments of Budget Statistics are responsible for the microcensus; the regular budget survey interviewers interview families belonging to the usual FBS sample, while families belonging to the supplementary sample are interviewed by staff specially recruited at the enterprises where their members work.

There were two distinct motivations for introducing the microcensus. Firstly it was a means of collecting information on questions which became foci of official concern in the 1950's and for which existing data sources were inadequate - in particular, information on fertility and on living conditions. Secondly it was intended as a source of reliable data on the distribution of family incomes. As we shall see (Chapters B3 and B4), the territorial and branch coverage of the FBS was and remains highly incomplete, and this is one of the reasons why its data are seriously biased. The microcensus covers all branches of the economy. (Originally agriculture was excluded, but since 1972 this too has been covered.) Therefore its data on income distribution are used to derive corrective coefficients for application to FBS data, in an attempt to compensate for the biases in the latter (see Chapter B9).

A major reason for the limited success of the attempt to use microcensus data to correct FBS data is that the microcensus, though more complete in its coverage than the FBS, is organised on similar principles and therefore shares many of the same deficiencies. Both surveys, for example, are subject to multi-worker bias and its associated biases with respect both to income and to family composition. This makes the microcensus not only a poor basis on which to construct corrective coefficients but also an unreliable source of fertility data.

At the statisticians' conference in 1968 (Vesesoyuznoe... 1969) there

was debate about the need for new surveys of fertility. Some said that there were enough data on fertility in the microcensus, but a researcher at the Scientific Centre for the Study of Population at Moscow University[1], G.P.Kiseleva, argued that microcensus data were unrepresentative, as families with low worker-dependent ratios were under-represented in the sample (p.299). From Table B1.8, based on data for Armenia in 1958, we can see that this is indeed one effect of the multi-worker bias, and that the ratio of number of working family members to total family size is subject to an upward bias of about 10 per cent.

Thus, though the conduct of supplementary surveys "on the budget network" is organisationally convenient, it has the unfortunate consequence that not only FBS data, but also the data from many other important surveys, are seriously unreliable on account of the faults in the FBS sample design. Study of the FBS therefore has a broader significance for the assessment of Soviet survey sampling in general.

Note to Appendix 1

1 Nauchnyi tsentr po izucheniyu narodonaseleniya MGU

APPENDIX 2 TO CHAPTER B1

DESIGN OF THE COLLECTIVE FARMER SAMPLE BEFORE 1952

Postnikov (1966) provides us with an account of the sample design for collective farmer families in the 1930's and 1940's. This account is not of purely historical interest, and provides important clues to the understanding of the contemporary sample design.

Collective farm agriculture in the pre-war period differed from collective farm agriculture from the 1950's onwards in several ways which affected the choice of the most appropriate sample design. First, collective farms at that time were much smaller than they became later: in 1932 the average size of a collective farm in the Central Non-Black-Earth Zone, the Upper Volga Zone and Belorussia was 39 households, and in the North-Western Zone just 25 households. Thus there were a large number of collective farms in each raion. Second, collective farms at that time did not own their own agricultural machinery but, if mechanised at all, were served by Machine-Tractor Stations (MTS). Whether or not a collective farm enjoyed access to the services of an MTS was a key factor in its situation. Third, payment for work on the collective farm was primarily in kind; money payment was much less prominent than it became from the time of Khrushchev onwards. Payment was also lower then, and less differentiated both among and within collective farms. Thus sown area per head was considered a better measure of the level of prosperity (or, rather, penury) of collective farms than payment per labour-day[1]; size of family was considered a better indicator of the situation of households than the number of labour-days worked, though whether or not a family possessed a cow was considered a significant indicator then as now.

Because of the extremely large number of collective farms per oblast', it was not convenient to select farms from oblast' lists directly. The selection process had accordingly not two but three stages: the selection of raiony within each oblast' covered by the survey; the selection of collective farms within each of the selected raiony; and the selection of families within each of the selected farms.

Each interviewer was responsible for collecting 36 budgets; presumably the amount of work needed to collect one budget was less than it became when the standard of living rose. Because of the very small size of many collective farms, only twelve budgets were collected on each collective farm surveyed. Each interviewer therefore covered a "budget point" consisting of three collective farms, which for practical reasons had to be in close proximity to one another in the same raion.

The procedure for selecting raiony within an oblast' covered by the survey (many oblasti were not) is illustrated in Table B1.11. The total number of collective farmer budgets allocated to the oblast' (180 in the illustration) was divided by the number of budgets to be collected per raion (36, by a single interviewer) to give the number of raiony to select (here 5). The names of the raiony in the oblast' were listed together with the number of collective farmer households resident in each one. The raion heading the list was selected at random, followed by the raiony situated in the ring surrounding the first raion, then by those in the next ring out and so on. Systematic selection was carried out from the implicit list of households, using the column of cumulative totals of numbers of households. The interval was calculated by dividing the total number of households in the oblast' by the number of raiony to be selected (85,000/5 = 17,000). The starting-point used was the end of the first interval, though the choice of starting-point is immaterial when the first raion is chosen at random; the method of concentric circles provides for a fairly even territorial

TABLE B1.11

PROCEDURE FOR THE SELECTION OF RAIONY IN THE PRE-WAR SURVEY OF COLLECTIVE FARMER BUDGETS

Name of raion	Number of households in the raion '000	Cumulative number of households '000	
A	3	3	
B	4	7	
C	2	9	
D	8	17	+
E	3	20	
F	2	22	
G	7	29	
H	6	35	+
I	7	42	
J	12	54	+
K	10	64	
L	11	75	+
M	10	85	+

Selected raiony marked +

Number of raiony in the oblast'	=	13
Number of households in the oblast'	=	85,000
Number of budgets to be collected in the oblast'	=	180
Number of budgets in one interviewer set	=	36
Number of raiony to be selected = 180/36	=	5
Interval (in households) = 85,000/5	=	17,000

Selection points: 17,000, 34,000, 51,000, 68,000, 85,000

Source: Postnikov (1966)

distribution of raiony whatever the starting-point.

The collective farms in each selected raion were classified (from 1936) into four groups:

I. Collective farms served by MTS

a. with sown area under harvest in the preceding year below average for the oblast';

b. with sown area under harvest in the preceding year above average for the oblast'

II. Collective farms not served by MTS

a. and b. as under I.

A collective farm was selected at random from among the farms belonging to the group predominating in the given raion. The two other collective farms to be covered by the interviewer working in that raion were the two nearest ones belonging to the same group. The necessity for the three collective farms to belong to the same group is not explained by Postnikov; it militates against both proportional representation of the groups within each raion and the convenience of the interviewer when the nearest collective farms to the one initially selected do not belong to the same group. We can only presume that the rule was designed to facilitate data-processing, at that time of course manual. Each interviewer presumably aggregated her 36 budgets, sending only summary totals of indicators in to the oblast' office, a procedure needed to keep down the volume of processing in the office. If breakdowns by group of collective farm were wanted - and the significance of the two classification variables, access to an MTS and sown area, makes it very likely that they were wanted - then each interviewer had to work at collective farms of a single type.

Did this method provide for proportional representation of the four groups of collective farms, as well as of collective farms of different productive orientations and "natural-historical zones", as claimed by Postnikov? If collective farms in one raion tended to have one set of characteristics

and those in another raion another set, then restricting the sample in each raion to the type of farm predominating in that raion should have had no serious ill effects. If, on the other hand, a given group of collective farms formed a substantial minority of the farms in most or all of the selected raiony but a majority in none of them[2], that group would not be represented at all in the sample, however important it might be for the agriculture of the oblast' as a whole.

From Table 8 in Postnikov (p.76) we can calculate typical numbers of raiony selected in each oblast' by dividing the sample sizes given by 36. Thus each oblast' covered in the Ukraine had 504 collective farmer budgets, implying the selection of 14 raiony. In some territories more raiony than this were selected, in others fewer (for example: Orenburg oblast' - 720 budgets, 20 raiony; the Mordovian ASSR - 216 budgets, only 6 raiony). In a territory with very heterogeneous agricultural conditions forming a complex territorial pattern, only a sufficiently large sample of raiony could be representative of the territory as a whole, even with a sound method of selection.

"The method of selection gave good results. Resampling was necessary extremely rarely, such cases being due to incorrect application of the method" (Postnikov). If territorial distributions of collective farms with different characteristics were such as to make the misgivings we have expressed groundless, this claim may well be true. This writer knows too little about Soviet agriculture of the period to judge the adequacy of the sample design to its structure.

Let us proceed to the selection of households within collective farms. All the households on a selected collective farm were classified into two groups - those with and those without a cow - and within each group by size of family. The sample for the collective farm was allocated proportionally among these groups and subgroups. The allocation must have been very

approximate, as only 12 families were to be selected on each collective farm, and it would have been sensible to take measures to ensure proportional representation by type of household at a level higher than that of the individual collective farm. Systematic selection was used to select households, presumably from lists for each subgroup of households.

Our doubts notwithstanding, the sample design as presented suggests that its creators had a sound grasp of sampling theory, especially considering that the 1930's were only the end of the formative period of sampling theory. The use of a cartographic technique to achieve territorial representativeness, and of a randomised starting-point in systematic sampling, contrast favourably with later practice. Moreover, the programme of representativeness checks included checks of the distributions of the most important variables as well as of the means of a wider range of variables[3], while checks of distributions were no longer conducted in the post-war survey.

Postnikov reports work carried out by the Sector of Collective Farmer Budgets (of the TsSU Department of Budget Statistics) in 1935-6 to assess sample representativeness with respect to those indicators available only from the budget survey. This work involved the calculation of theoretical sampling errors - and, in contrast to later work (see Chapter B6), actually used the correct formula for cluster sampling. As the work of calculation by hand was very laborious, it was done for a restricted number of survey variables, for one month's data only, and only for Moscow oblast'. In Table 7 (p.73), nevertheless, Postnikov presents relative sampling errors for eleven oblasti. This all confirms the sound grasp of mathematical statistics which survey statisticians still had at this period.

Notes to Appendix 2

1 Besides, it is an effort to estimate payment per labour-day on a standard measure when payment is mainly in the form of a variety of food products of different kinds and qualities.

2 "Majority" in the sense of the largest of all the component proportions, not necessarily an absolute majority.

3 Mean indicators of the selected collective farms were compared with corresponding means for all the collective farms of the oblast' on the following indicators: population; head of collective-farm livestock by types; sowings and yields of grain, potatoes, vegetables, sunflower, flax and hemp, and sugar-beet; amounts of grain, potatoes, vegetables and money paid for one labour-day. Checks were made on the distributions of the selected collective farms by the following indicators: distance of collective farm from the nearest town and from the nearest railway station; yield of grain and of potatoes per hectare; amounts of grain and of potatoes paid for one labour-day.

Mean indicators of the selected households on each collective farm were compared with means for all households on the following indicators: number of persons per family; number of labour-days worked per year per family; area of sowing on the private plot; number of privately owned livestock per family. Checks of distribution were made on private possessions of livestock (percentage of households without livestock, without large horned livestock, without cows, without pigs, without sheep or goats; percentage owning 1-2, 3-5, 6-10 and more than 10 head of sheep or goats).

Comparisons of mean indicators of the selected raiony with oblast' means were also conducted on (unspecified) basic indicators. Discrepancies of up to 5-6 per cent were considered tolerable.

CHAPTER B2

SIZE AND SOCIAL-GROUP STRUCTURE OF THE SAMPLE

1 Introduction

In this chapter we trace the size and the social-group structure of the FBS sample from 1928, when the survey was first organised on a centralised basis, up to 1981, concentrating on the period since the FBS was reconstructed in approximately its present shape in 1951-2. We also discuss the different attitudes expressed by Soviet writers regarding the desirable size of the FBS sample.

In Section 2 we explain how the Soviet population is classified into "social groups" for the purpose of organisation of the FBS. Section 3 describes the development of the sample size and of the social-group breakdown of the sample over time with the help of graphs. A detailed discussion of the sources used is relegated to the Appendix to this chapter, along with the figures on which the graphs are based and the explanation of the derivation of these figures[1].

Having assessed the size of the FBS sample by international standards, we proceed to consider the various Soviet views on the question of the sample size in Section 4. We sum up the main points in Section 5.

2 The classification of the Soviet population into "social groups"

The organisation of the FBS is based on the official classification of the adult population into such "social groups" (obshchestvennye gruppy) as "workers", "employees" and "collective farmers". Available data on the social-group structure of the FBS sample are therefore also based on this classification, the nature of which requires some explanation.

The main terms used in the classification and their inter-relations

are set out in Table B2.1. The adult population is first divided into three general categories:

(a) "Workers and employees" in the employment of State enterprises, establishments and organisations;

(b) "Collective farmers", members of collective farms; and

(c) Various groups of persons not in employment, such as students, old-age pensioners and the disabled.

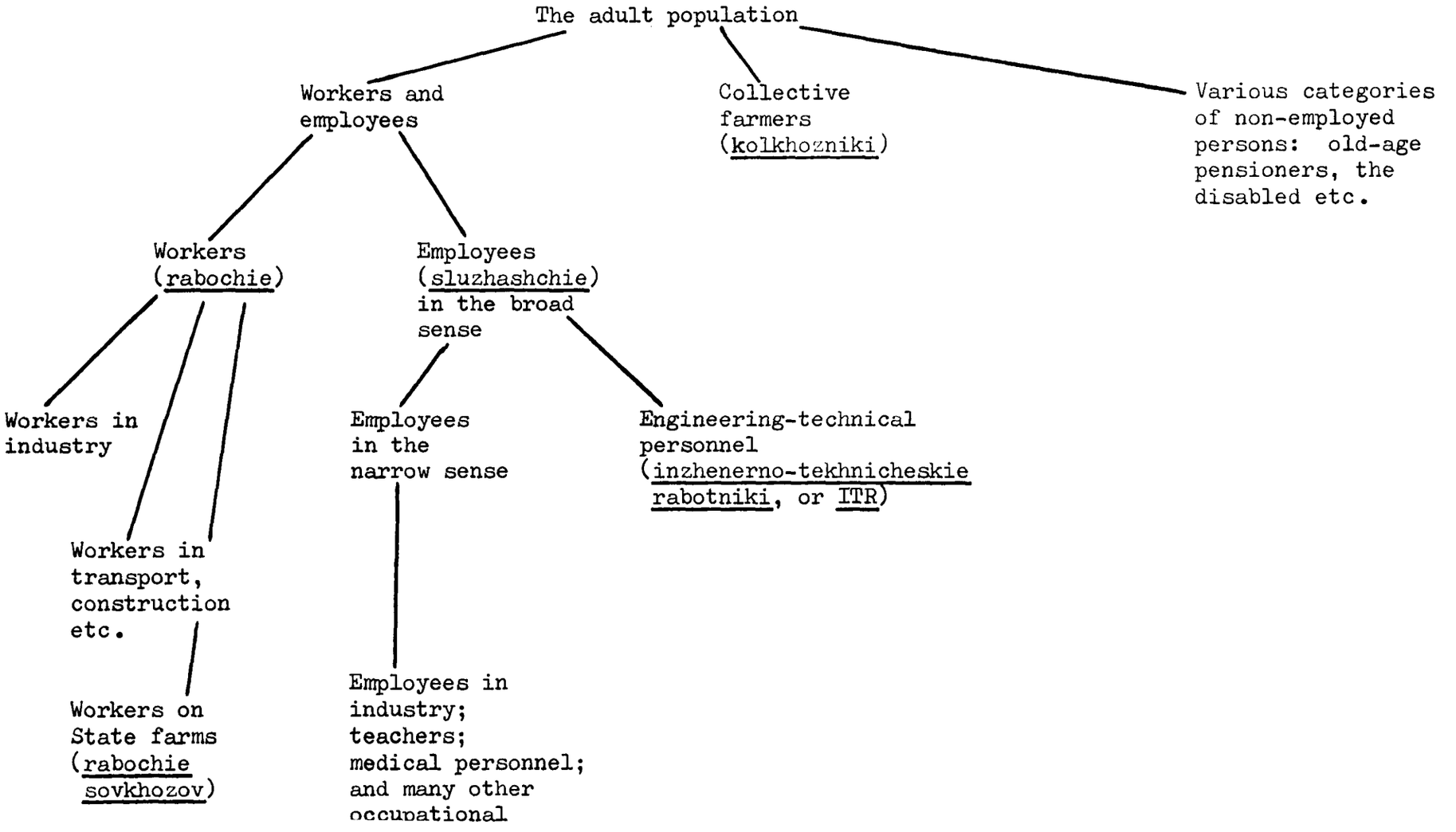
"Workers and employees" are divided into manual "workers" and non-manual "employees". "Workers" are further subdivided according to the branch of the national economy in which they are employed: industry, transport, construction, agriculture and so on. "Workers" in agriculture are mainly those employed on State farms.

"Employees" are subdivided into groups of employees employed in various productive branches of the national economy (employees in industry, employees in construction etc.) and many occupational groups (educational personnel, medical personnel, scientific personnel etc.). A distinction is sometimes made between "engineering-technical personnel" (ITR) and other employees: the term "employees" may be used to include ITR (employees in the broad sense) or to exclude them (employees in the narrow sense). We shall use the term in its broad sense (except in the Appendix to this chapter).

The official classification of the population into these "social groups" is an administrative, legal and ideological one. It does not correspond to any of the classifications of the population into "socio-economic groups" (sotsialno-ekonomicheskie grupry) used by Soviet social scientists for the purpose of substantive socio-economic analysis. Three important differences between the two types of classification should be noted:

(a) The classification into "social groups" is a classification of individuals by economic function, while classifications into "socio-economic

TABLE B2.1 THE OFFICIAL CLASSIFICATION OF THE SOVIET POPULATION INTO "SOCIAL GROUPS"



groups" are classifications of families by their conditions of life. In the FBS families are allocated to "social groups" - "families of workers", "families of employees" etc. - in accordance with the status of the family member selected for participation in the survey. As very many families contain members belonging to two or more "social groups", the allocation of families to "social groups" is often arbitrary.

(b) Classifications into "socio-economic groups" take account of the level of family income (or per-capita income), which is ignored by the classification into "social groups". The category "employees" covers the full range of income levels.

(c) Classifications into "socio-economic groups" usually distinguish between urban and rural families. The urban-rural division is overridden by the classification into "social groups". Members of collective farms belong to a different "social group" ("collective farmers") from workers and employees on State farms ("workers" and "employees"), because collective farms are formally collective group property while State farms are direct State property. However, the conditions of life of workers and employees on State farms resemble those of collective farmers much more closely than they do those of urban workers and employees[2].

The consequences of the organisation of the FBS on the basis of "social groups" rather than on that of "socio-economic groups" for the data produced from the survey are considered in Chapter C3.

In this chapter we shall make use of the following system of abbreviations. The FBS sample will be considered as the sum of five components:

"workers" exclusive of "workers" on State farms	w
"employees"	e
"workers" on State farms	s
"collective farmers" (kolkhozniki)	k
old-age pensioners	p.

Combinations of letters will be used to represent various combinations of components:

"workers" inclusive of "workers" on State farms	ws (or w+s)
"workers and employees"	we (or w+e)
those working on collective and on State farms	ks (or k+s)
etc.	

The groups (we) and (ks) are the closest possible approximations to the urban and rural subsamples respectively. The correspondence is not exact because it has not been possible to reallocate the few rural "workers and employees" in the sample who do not work on State farms, such as rural teachers and medical personnel, from the predominantly urban to the predominantly rural category.

3 Development of the size and social group structure of the FBS sample over time

The development over time of the size of the FBS sample and of its social-group structure is shown in the graphs. Table B2.2 shows the representation in the sample of each of the five categories (w, e, s, k, p), with alternative lines for "workers" exclusive and inclusive of those on State farms (w, ws). The following two graphs show the total sample size and its composition, accumulating components in different orders. In Table B2.3 "urban" workers are combined with employees to give the "urban" sample. In Table B2.4 workers on State farms are combined with collective farmers to give the "rural" sample, to which "urban" manual workers are added to give the full "manual" sample. The sample of old-age pensioners is indicated at the top of each graph.

Table B2.4 also provides a comparison of the size of the FBS sample with the sample sizes used in two Western budget surveys: the Family Expenditure Survey (FES) conducted in Britain by the Office of Population

TABLE B2.2 - USSR SAMPLE SIZES BY SOCIAL GROUP

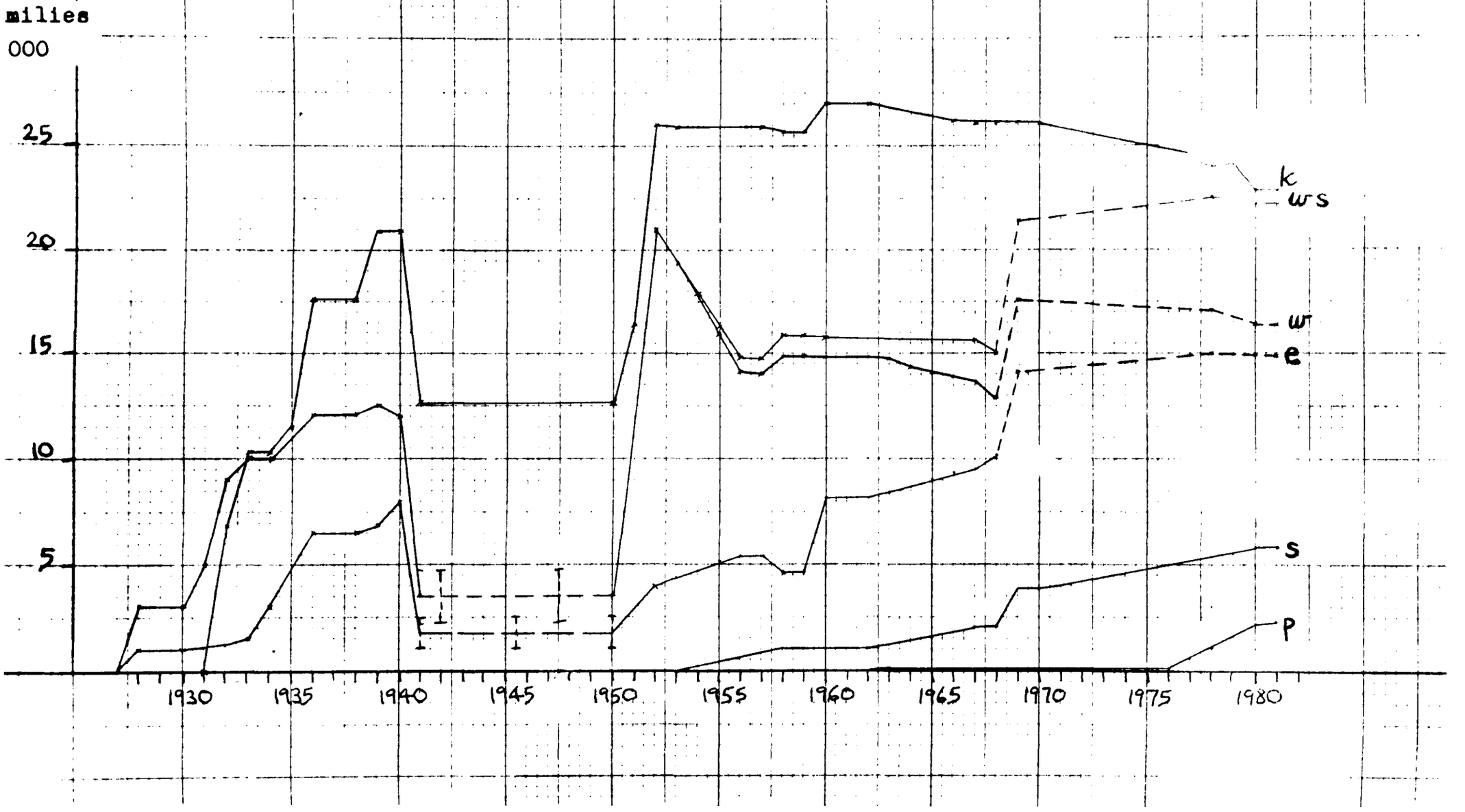


TABLE B2.3 - USSR SAMPLE SIZES BY SOCIAL GROUP (CUMULATIVE GRAPH A)

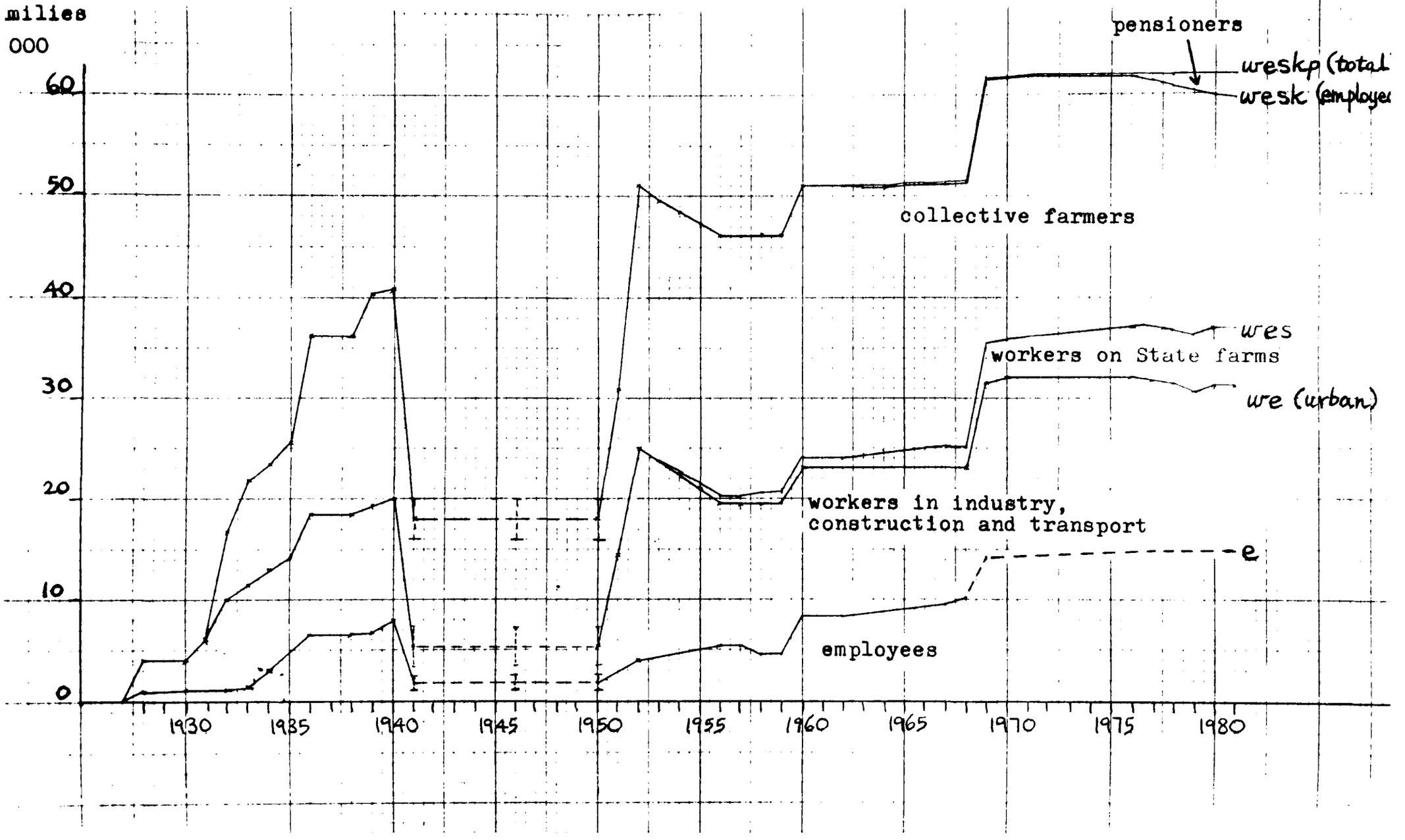
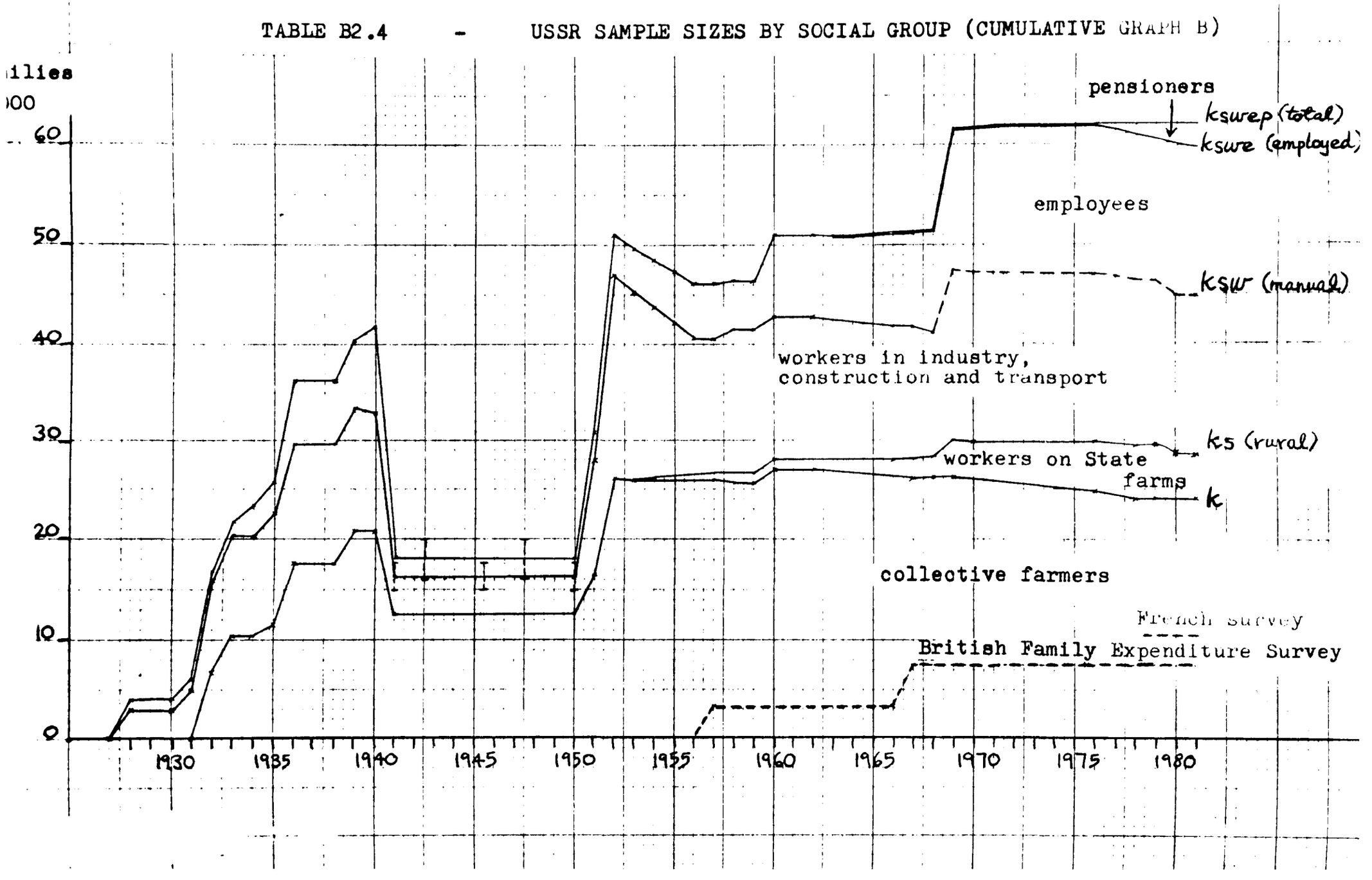


TABLE B2.4 - USSR SAMPLE SIZES BY SOCIAL GROUP (CUMULATIVE GRAPH B)



Censuses and Surveys, and the budget survey conducted in France by the Institut National de la Statistique et des Etudes Economiques (INSEE).

While Tables B2.2 - B2.4 show absolute sample sizes in thousands of families, Table B2.5 shows sampling fractions in families sampled per 1000 families of the general population. Urban, rural and overall sampling fractions are shown separately. The figures on which Table B2.5 is based are presented in Table B2.6.

In the 1920's budget surveys of local scope were conducted by a variety of organisations, such as trade unions, as well as by TsSU. A survey of the family budgets of workers and employees at national level was set up by TsSU in 1928, and by 1929 budget survey work had been completely transferred into the hands of TsSU (Matyukha 1969). Following the collectivisation of agriculture, the survey of the family budgets of collective farmers began in 1932.

Coverage of workers, of employees and of collective farmers expanded fairly steadily throughout the 1930's, and the survey of collective farmers was gradually extended to cover new territories (Postnikov 1966). The budget survey was drastically curtailed by the Second World War, and remained in its wartime state until 1951.

In 1951-2 the survey was reconstructed in a form which has not changed in any fundamental way since then. The principles of its reconstruction were codified in the only decree of the Council of Ministers known to have been promulgated on the survey, dated November 1951 (Resheniya... 1968 pp.670-3). The decree laid down that 51,000 families were to be surveyed (about 0.1 per cent of the population of the USSR): 21,000 workers, 4,000 employees and 26,000 collective farmers. These sample sizes were achieved during 1952. They provided for a coverage of the urban population slightly more intensive than that of the rural population (1.1 per 1,000 as against 0.9 per 1,000).

TABLE B2.5 - USSR SAMPLING FRACTIONS PER 1000 FAMILIES OF POPULATION

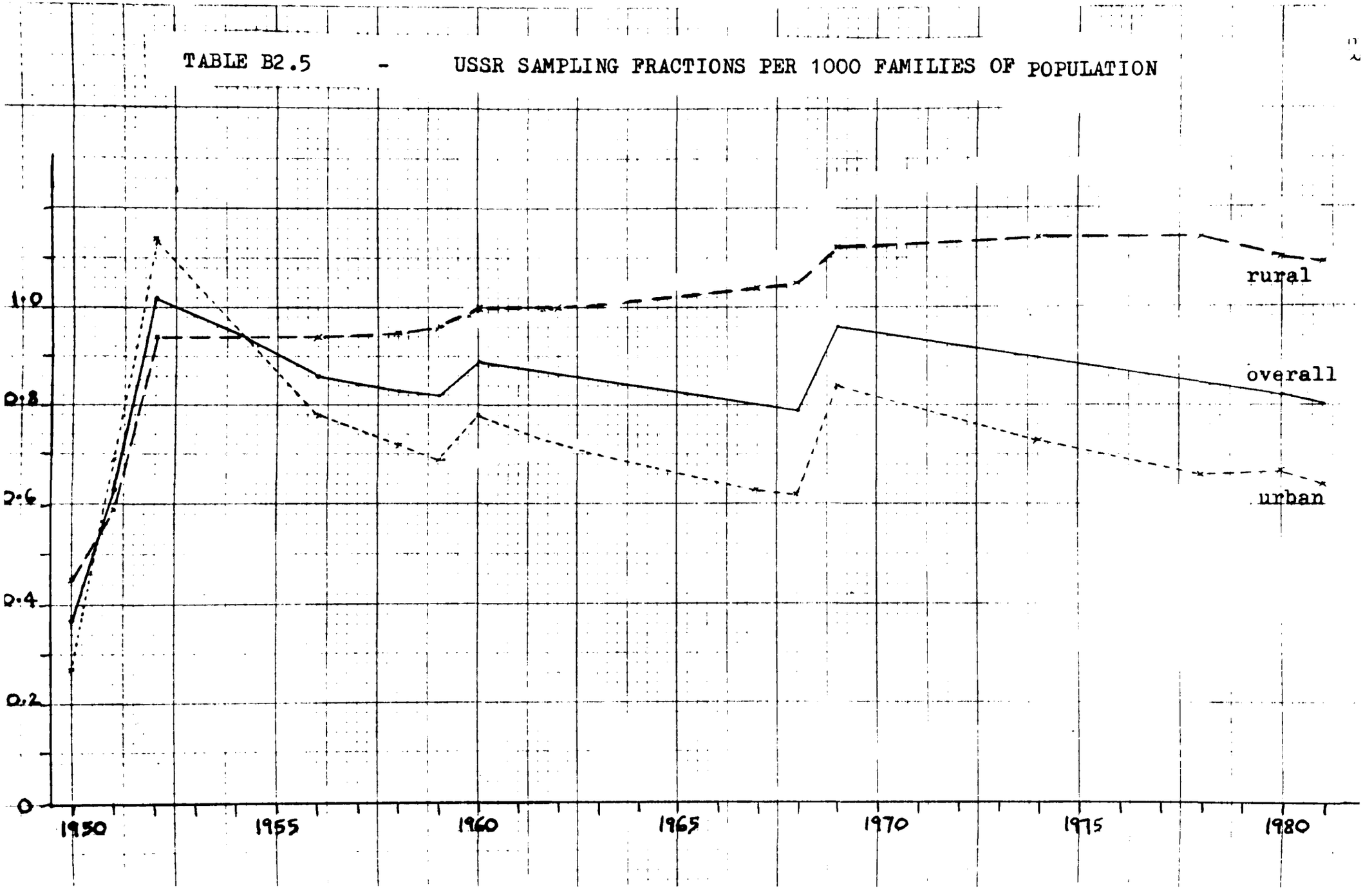


TABLE B2.6

USSR SAMPLING FRACTIONS

	Budgets per 1000 families		
	Urban population	Rural population	Overall
1950	0.3 ($\pm$0.1)	0.5	0.3 - 0.4
1951	0.7	0.6	0.6
1952	1.1	0.9	1.0
1953	1.0	0.9	1.0
1954	1.0	0.9	1.0
1955	0.9	0.9	0.9
1956	0.8	0.9	0.9
1957	0.8	0.9	0.8
1958	0.7	1.0	0.8
1959	0.7	1.0	0.8
1960	0.8	1.0	0.9
1961	0.7	1.0	0.9
1962	0.7	1.0	0.9
1963	0.7	1.0	0.9
1964	0.7	1.0	0.8
1965	0.7	1.0	0.8
1966	0.6	1.0	0.8
1967	0.6	1.0	0.8
1968	0.6	1.0	0.8
1969	0.8	1.1	1.0
1970	0.8	1.1	0.9
1971	0.8	1.1	0.9
1972	0.8	1.1	0.9
1973	0.8	1.1	0.9
1974	0.7	1.1	0.9
1975	0.7	1.1	0.9
1976	0.7	1.1	0.9
1977	0.7	1.1	0.9
1978	0.7	1.1	0.8
1979	0.7	1.1	0.8
1980	0.7	1.1	0.8
1981	0.6	1.1	0.8

Method of calculation: For urban, rural and overall sample sizes see Table B2.7, columns we, ks and weskp respectively (but here we do not exclude pensioners from the urban and rural samples). These are divided into population figures, taken from the Nar SSSR statistical handbooks and then converted into numbers of families using the average USSR family sizes given in Table B3.8. Considerable imprecision arises from the interpolation and extrapolation of family sizes and from the assumptions used in estimating sample sizes (see appendix to this chapter). No adjustments have been made to transfer rural teachers and medical staff from the "urban" we to the "rural" ks sample (but there are not many of these). We therefore record the estimated sampling fractions to one significant digit only.

It seems - though this is not openly stated in our sources - that there was some difficulty in maintaining the 1952 sample sizes, for in 1956-7 the sample was only 46,000 strong and it was restored to 51,000 only in 1960. For 1958-9 we have an indication for the first time of the inclusion of a specified number (1,000) of families of workers on State farms.

The contraction of the sample betwen 1952 and 1956 was restricted to workers; the employee sample continued to expand during this period, while the size of the collective farmer sample was stable. Whatever the reason for this phenomenon, the effect was to invert the relative intensity of rural and urban coverage; by 1956 the rural sampling fraction was still 0.9 while the urban sampling fraction had fallen to 0.8 per 1,000. The gap continued to widen as the urbanisation of the population proceeded, and was only partly closed in 1960, when only 4,000 of the 5,000 lost workers and employees were replaced while the number of collective farmers rose to 27,000.

The size and composition of the sample remained about the same between 1960 and 1968, except that employees - there were about 10,000 employees' families by 1968 - and perhaps State farm workers gained at the expense of the other groups. By 1968, as a result of the failure to adjust sample allocation to the declining relative weight of the rural population, the urban sampling fraction (0.6 per 1,000) was not much more than half the rural (still 1.0).

In 1966 the Government discussed how to improve the budget survey (Starovskii 1969), but it took until 1969 for TsSU to make changes in the survey in response to this discussion. The total sample size was increased by 10,500 to 62,000, at which level it still remains. The whole of the increase consisted of workers and employees (from 25,100 to 35,600), though this was enough only partly to close the gap between urban and rural

sampling fractions (bringing the former up from 0.6 to 0.8 per 1,000). The branch structure of the worker sample was improved, with significant increases in the numbers of workers outside industry - that is, in construction, on the railways and on State farms. The coverage of some branches, such as food industry and residential services, was introduced for the first time, and coverage of doctors, nurses and teachers was significantly expanded (Soveshchanie... 1969). Starovskii (1969) informs us that some corrections were also being made to the territorial distribution of the collective farmer sample, but we have no specific information on this.

No further substantial reorganisation of the sample has been carried out since 1970. The number of workers (exclusive) and of employees have remained about the same. The collective farmer sample has at last, in response to the long decline in the collective farmer population, fallen below its 1952 level to 22,700 families in 1981, while the number of State farm workers has risen to 5,800; collective farmers are still, however, much more intensively represented than State farm workers. The most important change of the last few years has been the formation for the first time of a substantial sample of old-age pensioners (5,800 families in 1980-1). The structure of the sample lags as far as ever behind the urbanisation of the population, with the rural sampling fraction in 1981 almost double the urban (1.1 per 1,000 as against 0.6).

The proportion of the population comprised by the sample has tended to follow a cyclical pattern since 1950: expansions of the sample (in 1951-2, 1960 and 1969) raise the overall sampling fraction above 0.8 per 1,000 (to 1.0, 0.9, 1.0 respectively), and in the intervening periods it gradually falls back again to 0.8 per 1,000 as the population rises while sample size remains constant. If past experience is any guide, we may expect a new expansion of the sample soon.

Even without any new expansion, however, the Soviet FBS remains extremely large by international standards - the largest in the world, as Soviet authors like to proclaim. Comparison with the size of the corresponding British survey brings this point home. The British Family Expenditure Survey at present covers 7,500 households, less than one-eighth the size of the Soviet sample (and we suspect it may soon be reduced in size). This direct comparison of sample sizes, without reference to population, has some relevance inasmuch as the sampling error of global estimates depends primarily on absolute sample size. But even a comparison of sampling fractions reveals that the USSR is almost twice as intensively covered by its budget survey as Britain is by the FES[3].

4 Soviet attitudes on the question of the size of the FBS sample

Large as the FBS sample is by international standards, TsSU has come under continuing pressure to expand the sample further from scholars who assume that this is necessary to improve sample representativeness:

> In order to enhance the role of the FBS in planning, it is necessary significantly to increase its size to cover all branches of the national economy (Sapelnikov, Shnirlin and Levin 1959).

Karapetyan (1980 p.127) makes the observation that

> at meetings on these problems, recommendations are usually made to increase the sample size in order to secure representative data. This is an error. Sample size is not the issue: it is the type of survey which prevents representativeness.

Those who take this point of view emphasise that increases in sample size do not in themselves improve representativeness, and that representativeness can be improved without increasing sample size:

> In the last eighteen years the number of family budgets surveyed has increased almost fourfold. But what has changed in essence? (Rimashevskaya 1971)
>
> Reconstruction of the FBS sample need not involve any significant increase in sample size. It is necessary

> only that the coverage of different population groups in a given territory take account of income levels, the age structure of families and other factors which influence demand (Lakhman and Frenkel 1967).

The assumption that improvements in representativeness require increase in the sample size appears to be shared by TsSU officials responsible for the FBS, who resist pressure to increase the sample size by arguing that such improvements would not be worth the cost entailed:

> (To cover all population groups) an unjustified increase in sample size would be required, involving significant additional expenditure on collecting and processing data (Matyukha 1967).

Considerations of representativeness apart, the general Soviet preference for large samples is also in evidence (see Chapter A4). An American marketing delegation, sent on a visit to the USSR in 1960 by the Harvard Graduate School of Business Administration, were told at the Ukraine Scientific Research Institute of Trade and Public Catering in Kiev that the institute had analysed data from more than 40,000 family budgets by occupation, age group, income level and so on (Edwards 1960). The Americans were unimpressed:

> When we told them that this figure (40,000) is extremely high and that we would use a smaller statistically selected sample of representative families, they said that this method of selection is not used in the Soviet Union, declaring that the more budgets studied, the better the survey. [4]

There have been, however, Soviet assessments of the optimal sample size conducted on a more scientific basis. These have sometimes concluded that a larger sample is needed, and sometimes that a smaller sample would suffice:

> Preliminary calculations of sampling errors by Chomaryan (1962) show that the reconstructed FBS must be at least five times larger... but some consider that a sample of 1,000 would be adequate. This would require finding stable strata representing the heterogeneity of the USSR and knowing their proportions in the population (Rimashevskaya 1968b).

Such discrepant recommendations are no doubt explained by variation in the levels of precision considered necessary by different investigators. In particular, an investigator who is concerned only with the precision of estimates relating to large regions and to the USSR as a whole will naturally be satisfied with a much smaller sample than an investigator who is concerned also with the precision of estimates at a more local level.

It is commonly held that "local economic calculations require increased sample sizes for individual Republics, krai, oblasti and towns" (Shvyrkov 1965). The demand to be able to use FBS data for economic calculations at the oblast' level will be further discussed in Chapters D2 and D5. Here we observe that, while this demand can be satisfied only by a large national sample, it nevertheless tends to take an exaggerated form as a result of distrust of samples of moderate size. For example, in the course of a discussion on methods of studying the private plot economy, at the conference of statisticians in 1968, Ye.I.Sukhova expressed the view that an oblast' subsample of 1,200 budgets is too small to be of any use:

> Some comrades say that data... can be obtained from the FBS. The sample for Moscow oblast' is 1,200 budgets, which cannot give a true picture; one cannot use it for practical purposes (Vsesoyuznoe... 1969 p.395).

While ignorance of sampling theory can lead to calls for an extremely large sample, in certain cases it can lead to opposite calls for a much smaller sample than the existing one. This takes place when the analyst adheres to the concepts of typological monography, as in the works of such specialists in budget studies as P.P.Maslov (1971 pp.35-6), Ye.O.Kabo (1972) and A.E.Khidirkhanova (1982). Maslov argues that complaints concerning the "quantitative representativeness" of the FBS are misguided, because the large number of "types" of family budget covered ensures its "qualitative representativeness". In fact, only a few budgets belonging to each

homogeneous "type" need be included:

> ... It is necessary to reduce our expensive budget sample. It is very easy to establish that a severalfold reduction would give basically the same results, because one may not require territorial representativeness from budget statistics. Its tasks are to reveal group differences - that is, differences among types of families. Such types may be identified on the basis of a small number of observations, because intra-type variability is (we repeat) here negligible.

Maslov recommends that the sample be reduced to one-quarter of its present size[5].

Maslov clearly considers existing knowledge of "types" of family budgets adequate for the necessary re-design of the FBS sample. Zhidikhanova also advocates "using types to select the sample instead of doing it mechanically" and thereby achieving a much smaller sample (Zhidikhanova 1982 pp. 119, 122), but she believes that intensive new research is needed to determine family "types":

> A theoretical typology of families will require significant expense. However,this expense will be recouped at a later stage..., inasmuch as the number of units surveyed will be reduced at the same time as qualitative representativeness will be achieved (p.123)[6].

Soviet attitudes concerning the optimal size of the FBS sample may then be determined either by proper sampling-theoretic calculations or by ideological assumptions. Neither approach rules out any particular conclusion. If sampling theory is used, the recommended sample size depends on the requirements made on the use of FBS data at different geographical levels. Ideological assumptions often give rise to calls for a larger sample. if the assumptions are those founded in the ideal of complete enumeration, but alternatively they may lead to advocacy of a much smaller sample. if the assumptions happen to be those deriving from the monographic tradition.

5 Conclusion

The FBS sample is a very large one by international standards. At long intervals the sample undergoes substantial expansion, and on these occasions the opportunity is taken to improve the coverage and social-group structure of the sample. This does not, however, suffice to prevent the structure of the sample lagging behind changes in the structure of the population - for example, with regard to urbanisation.

Soviet attitudes regarding the optimal size of the FBS sample are very varied, and are influenced both by practical requirements and by ideological assumptions.

Notes to Chapter B2

1 No available source systematically sets out the information presented in this chapter, which has been collated from numerous disparate sources, some of which are unclear or contradict one another. Gaps in the available information have been filled in to assemble an overall picture using assumptions which in some cases are rather arbitrary. While we are confident that the broad outline of the picture is correct, some of the estimates are no doubt significantly in error.

2 The inhabitants of rural areas suffer from the restricted range of goods and services available in those areas, irrespective of their "social group". Unlike most urban inhabitants, they generally have private plots. The introduction of guaranteed minimum incomes for collective farmers reduced the former difference between the systems of remuneration operating on collective and on State farms.

3 The population of the USSR is about 4.5 times as great as that of Britain, while the Soviet FBS is almost nine times larger than the British FES.

The sample of the French budget survey - 10,600 families in 1978-9 (Glaude 1982) - is rather larger than that of the British, and the population of France is somewhat smaller than that of Britain. France is therefore intermediate between Britain and the USSR in intensity of coverage.

4 I.I.Korzhenevskii, an economist at the same institute, asked Dr.P.Hanson, on a visit in 1965, how Britain and other Western countries could possible manage with budget surveys having such small samples.

5 Maslov, it is true, mentions that "in 1969 experimental calculations conducted under the leadership of A.Kormishevaya confirmed the possibility of reducing the FBS sample".
These calculations were presumably based on sampling theory. He also refers to British work on Family Expenditure Survey data. But his advocacy of a small sample is philosophically predetermined; the existence of calculations which seem to support his point of view is just grist to the mill.

6 Zhidikhanova throws little light on the methodology of her proposed "theoretical typology". She makes no reference to recent Soviet work concerned with constructing a typology of consumer budgets by means of mathematical-statistical analysis of budget data, carried out at the Central Economic-Mathematical Institute (Tipologiya... 1978).

For further discussion of the relationship between the monographic tradition and the FBS, see Chapter B12.

APPENDIX TO CHAPTER B2

DERIVATION OF SAMPLE SIZE DATA FROM SOVIET SOURCES

A. THE PRE-WAR SURVEY 1928-1941

Between 1928 and 1931 the survey covered workers and employees (excluding ITR) only. The survey of collective farmers began in 1932. ITR were covered from 1934.

Krapivina (1966) gives:

1928-30	3000 w + 1000 e	=	4000 we	
1932	9000 w + "more than 1000" e	=	more than 10000 we	
1933	10000 w + 1500 e	=	11500 we	
1934	10000 w + 1500 e + 1500 ITR	=	13000 we	
1936-8	"more than 12000" w + 5000 e + "more than 1500" ITR	=	18500 we	
1939	12500 w + 5185 e + 1675 ITR	=	19360 we.	

These figures are fairly consistent with those of Matyukha, Postnikov and Samoilov (1958):

1931	5000 w		
1935	14200 we		
1940-1	12000 w + 8000 e	=	20000 we.

Yezhov (1965 p.316) gives exactly the same figures. Uchebnoe... (1958) gives 19200 we "before the war". Matyukha (1969 p.420) gives:

1934	10000 w + 3000 e/ITR	=	13000 we
1939	12500 w + 7500 e/ITR	=	20000 we.

The figures given by Kats (1960 p.168) are difficult to reconcile with those from other sources:

November 1928 - September 1930 1500 w + 1500 e = 3000 we
October 1930 7500 w + 3000 e = 10500 we.

We take collective farmer sample sizes from Postnikov (1966):

1932 6800 k
1933-4 10300 k
1935 11500 k
1936-8 17700 k
1939 20924 k,

and from Matyukha, Postnikov and Samoilov (1958):

1941 21000 k.

We assume 20924 k for 1940 also, in line with the figure of 21000 k "before the war" given by Uchebnoe... (1958). Matyukha (1969 p.425) gives:

1932 6500 k
1938 about 17000 k
1940 21000 k.

The figures for 1932-4 are, however, subject to doubt, especially that for 1933, as Yezhov (1965 p.313) gives:

1932-3 5740 k
1934-6 9700 k
1937-9 17000 k
1940-1 21000 k.

Similarly, Dolgushevskii and Khristich (1976 p.339) give:

1932 6500 k
1933 5700 k
1934 9700 k.

Thus we have the following total sample sizes for the pre-war period (figures subject to serious doubt marked ?):

Year	w+e+ITR	k	total
1928-30	4 000 ?	-	4 000 ?
1931	6 000 ?	-	6 000 ?
1932	10 000	6 800	16 800
1933	11 500	10 300 ?	21 800 ?
1934	13 000	10 300	23 300
1935	14 200	11 500	25 700
1936-8	18 500	17 700	36 200

1939	19 360	20 924	40 284
1940	20 000	20 924	40 924
1941	20 000	21 000	41 000

B. THE SURVEY DURING AND IMMEDIATELY AFTER THE WAR 1941-50

Following the Nazi invasion in 1941 the survey was severely contracted to cover only some Eastern regions and then remained at the same level until its reconstruction in 1951-2. Both Postnikov (1966) and Matyukha, Postnikov and Samoilov (1958) give a sample of 12,700 k in this period. We have no direct reference for urban and total sample sizes, but can derive them indirectly with a wide margin of error.

Matyukha (1966) states that the total sample was expanded 2.5 times in 1951-2; given the post-expansion sample of 51,000, this implies a pre-expansion sample of about 20,000, that is, 7,300 w+e+ITR + 12,700 k. Rimashevskaya (1971), on the other hand, refers to an almost fourfold expansion of the sample over the period 1951-1969; given a sample at the end of 1969 of 61,800, this implies a sample in the region of 16,000 in 1950, that is, 3,300 w+e+ITR + 12,700 k. As a compromise between the results of these two calculations, we use 5,300 w+e+ITR + 12,700 k = 18,000, with the figures for urban and total samples subject to error of ± 2,000.

C. THE POST-WAR SURVEY 1951-81

Expansion of the sample began in 1951; by the time of the decree on the budget survey of 3 November 1951 (Resheniya... 1968 pp. 71-3, it had reached 14,400 w+e + 16,400 k = 30,800. Expansion continued in 1952 to 21,000 w + 4,000 e + 26,000 k = 51,000 (Matyukha 1966). By 1956-7, however, the sample was only 14,800 w + 5,400 e + 25,900 k = 46,100

(Matyukha, Postnikov and Samoilov 1958). In 1958-9 the sample remained near this level, and we have a figure for workers on State farms (s) for the first time: 14,900 w + 4,700 e + 1,000 s + 25,700 k = 46,300 (Uchebnoe... 1958; Kats 1960 p.173 for 1959).

No source makes any explicit reference to this reduction in the sample of workers, but it appears that the 1952 sample size was restored in 1960: Matyukha (1960) gives 51,000 in all, and Starovskii (1962) gives 24,000 w+e 27,000 k = 51,000. Krosnoshchekov (1962) confirms that the sample was increased by 5000 in 1960.

Between 1960 and the beginning of 1969 the evidence points to no great changes in the sample:

1960			51 000 (Matyukha 1960
1962	24 000 w+e 27 000 k	=	51 000 (Starovskii 1962)
1964			51 000 (Krasnoshchekov 1964)
1965	25 000 w+e 26 000 k		51 000 (Yezhov 1965 p.316)
1965	29 000 w+e		(Rimashevskaya 1965)
1967	15 700 w 26 100 k		(Matyukha 1967)
1968	15 000 w		(Rabochii.. 1969 p.230)
	25 200 w+e 26 200 k	=	51 400 (Venetskii and Matyukha 1968)
1969 (start)	25 100 w+e 26 200 k	=	51 300 (Korovkin 1969)

The high figure given by Rimashevskaya for 1965 is inconsistent with the other references; it is more likely to be an error than the only indication of a substantial expansion and subsequent re-contraction of the sample in the mid-1960s, and so we leave it out of account.

In 1969 the sample of workers and employees was increased by 10,500, giving 35,600 w+e + 26,200 k = 61,800 at the end of the year (Korovkin 1969; Soveshchanie... 1969). However, Matyukha (1969) dates this expansion in 1968. The NKh SSSR statistical handbooks for 1973-77 give a

total sample size of 62,000, and Maslov, Romanovskaya and Simchera (1978) give 37,500 w+e + 24,400 k = 61 900. Posobie... (1980 p.373) and Dumnov, Pisarenko and Te (1981) both give the present sample size as 62,000, as well as the following sample breakdowns:

	1980	1981	
w + e	31 200	31 200	
s	5 800	5 800	
w + e + s	37 000	37 000	(by addition)
k	22 800	22 700	(by subtraction)
w + e + s + k	59 800	59 700	(by subtraction)
pensioners	2 200	2 300	
total	62 000	62 000	

This reflects the growth in the samples of State farm workers and of old-age pensioners over the last few years.

In order to construct the graphs showing sample sizes by social group, it was necessary to estimate sample sizes for those group-year combinations for which we lack clear information. A number of assumptions were required to do this, as well as time interpolations, and their arbitrariness makes the graphs only very approximately reliable.

The assumptions used were:

(a) After 1968 we have no sample sizes for workers or for employees alone, only for the two combined. We assume that workers continue to constitute 60 per cent of the combined sample. This is such a gross assumption that the line segments affected by it are shown as broken in the graphs.

(b) In rough agreement with the evidence on coverage of pensioners we assume that they were covered for the first time in 1963, that 100 were surveyed each year until 1976, and that the sample then grew steadily to reach the known figure of 2,200 in 1980. We assume that our sources included the pensioner sample under "workers and employees" until 1976,

and that in 1977-9 it was included partly under "workers and employees" and partly under "collective farmers" in porportion to the sample sizes of these groups.

(c) Similarly, we assume that workers on State farms were first covered in 1954 (200) rising steadily to the known figure of 1,000 in 1958-9, continuing the rise from 1962 to the known figure of 5,800 in 1980-1. We know that the number of workers on State farms was increased substantially in 1969: we assume from 2,100 to 3,800. We also assume that, where a reference does not make explicit the exclusion of workers on State farms (as well as construction and transport workers) from its figure for "workers" by using the term "workers of industry", "workers" includes workers on State farms. We then subtract our estimate for State farm workers to get the estimated number of non-agricultural workers.

Table B2.7 shows the figures on which the graphs are based. Known figures are marked by asterisks; all other figures are estimates.

TABLE B2.7 SAMPLE SIZE FIGURES USED IN GRAPHS

Year	p	s	e	w	ws	k	ks	ksw
1928-30	0	0	1000*	3000*	3000	0*	0	3000
1931	0	0	1000	5000*	5000	0*	0	5000
1932	0	0	1000*	9000*	9000	6800*	6800	15800
1933	0	0	1500*	10000*	10000	10300*	10300	20300
1934	0	0	3000*	10000*	10000	10300*	10300	20300
1935	0	0	4750	11000	11000	11500*	11500	22500
1936-38	0	0	6500*	12000*	12000	17700*	17700	29700
1939	0	0	6860*	12500*	12500	20924*	20924	33424
1940	0	0	8000*	12000*	12000	20924*	20924	32924
1941-50	0	0	1750 ($\pm$750)	3500 ($\pm$1250)	3500 ($\pm$1250)	12700*	12700	16200 ($\pm$1250)
1950 end	0	0	2800	11600	11600	16400*	16400	28000
1952 end	0	0	4000*	21000*	21000	26000*	26000	47000
1953	0	0	4400	19400	19400	25900	25900	45300
1954	0	200	4700	17700	17900	25900	26100	43800
1955	0	400	5100	15900	16300	25900	26300	42200
1956	0	600	5400*	14200	14800*	25900*	26500	40700
1957	0	800	5400	14000	14800	25900*	26700	40700
1958-59	0	1000*	4700*	14900*	15900*	25700*	26700	41600
1960-61	0	1000	8200	14800	15800	27000	28000	42800
1962	0	1000	8200	14800	15800	27000*	28000	42800
1963	100	1200	8400	14600	15800	26800	28000	42600
1964	100	1400	8700	14300	15700	26600	28000	42300
1965	100	1600	9000	14100	15700	26400	28000	42100
1966	100	1800	9300	13900	15700	26200	28000	41900
1967	100	2000	9500	13700	15700*	26100*	28100	41800
1968	100	2100	10200*	12900	15000*	26200*	28300	41200
1969								
start	100	2100	10200	12800	14900	26200	28300	41100
end	100	3800	14200	17600	21400	26200	30000	47600
1970	100	3800	14300	17800	21600	26000	29800	47600
1971	100	4000	14400	17800	21800	25800	29800	47600
1972	100	4200	14500	17700	21900	25600	29800	47500
1973	100	4400	14600	17600	22000	25400	29800	47400
1974	100	4600	14700	17500	22100	25200	29800	47300
1975	100	4800	14800	17400	22200	25000	29800	47200
1976	100	5000	14900	17300	22300	24800	29800	47100
1977	600	5200	15000	17200	22400	24400	29600	46800
1978	1100	5400	15000	17100	22500	24000	29400	46500
1979	1600	5600	14900	16800	22400	24100	29700	46500
1980	2200*	5800*	14800	16400	22200	22800*	28600	45000
1981	2300*	5800*	14800	16400	22200	22700*	28500	44900

Columns:

p	old-age pensioners
s	workers on State farms
e	employees (including ITR)
w	manual workers, excluding those on State farms
ws	manual workers, including those on State farms
k	collective farmers
ks	rural residents (collective farmers and State farm workers)
ksw	rural and urban manual workforce (all except employees and pensioners)

cont'd

Table B2.7 (cont'd)

Year	we	wes	wesk/kswe	weskp/kswep
1928-30	4000*	4000	4000	4000
1931	6000	6000	6000	6000
1932	10000*	10000	16800*	16800
1933	11500*	11500	21800*	21800
1934	13000*	13000	23300*	23300
1935	14200*	14200	25700*	25700
1936-38	18500*	18500	36200*	36200
1939	19360*	19360	40284*	40284
1940	20000*	20000	40924*	40924
1941-50	5300 (±2000)	5300 (±2000)	18000 (±2000)	18000 (±2000)
1951 end	14400*	14400	30800*	30800
1952 end	25000*	25000	51000*	51000
1953	23800	23800	49700	49700
1954	22400	22600	48500	48500
1955	21000	21400	47300	47300
1956	19600	20200*	46100*	46100
1957	19400	20200*	46100*	46100
1958-59	19600*	20600*	46300*	46300
1960-61	23000	24000	51000*	51000
1962	23000	24000*	51000*	51000
1963	22900	24100	50900	51000
1964	22900	24300	50900	51000*
1965	23000	24600	51000	51100
1966	23100	24900	51100	51200
1967	23100	25100	51200	51300
1968	23000	25100	51300	51400
1969				
start	22900	25000	51200	51300*
end	31700	35500	61700	61800*
1970	32000	35800	61800	61900
1971	32100	36100	61900	62000
1972	32100	36300	61900	62000
1973	32100	36500	61900	62000*
1974	32100	36700	61900	62000*
1975	32100	36900	61900	62000*
1976	32100	37100	61900	62000*
1977	31800	37000	61400	62000*
1978	31400	36800	60800	61900*
1979	30700	36300	60400	62000
1980	31200*	37000*	59800*	62000*
1981	31200*	37000*	59700*	62000*

Columns:

we	urban residents (workers, excluding those on State farms, and employees)
wes	all except collective farmers and pensioners
wesk	all except pensioners
weskp	total sample - all including pensioners

CHAPTER B3

TERRITORIAL COVERAGE OF THE SAMPLE

1 Introduction

In Chapter B1 it was noted that in principle the territorial distribution of the sample over the USSR is supposed to be proportional to population. In this chapter we examine the extent to which this principle is applied in practice. The validity of the claims of Soviet critics of the FBS (for example, Levin 1973 p.254) to the effect that territorial coverage is very partial and uneven will be confirmed. To the limited extent that available information permits, we shall analyse the way the sample is distributed territorially and consider possible rationales underlying the pattern of distribution.

The collection and processing of FBS data are normally organised by statistical offices at oblast' level - that is, by those of oblasti, krai, ASSR's and Union Republics without oblast'-level subdivision. As of 1978, there were 154 territories at this level in the USSR[1]. In Section 2 we consider how many of these territories are covered by the FBS. Where possible we also estimate how many territories are covered in individual Union Republics.

In Section 3 we attempt to establish the identity of the territories covered by the survey. Available information enables us to identify somewhat over one-half of the territories covered. There is no reason to suppose that the identified territories constitute a representative sample of the population of covered territories. Our discussion in Section 4 of the overall range of geographical coverage of the FBS, and of the criteria according to which territories may have been selected for coverage, is

therefore incomplete and tentative. Only the appearance of much fuller information about the FBS will make it possible to draw clear conclusions.

In Section 5 we analyse sample sizes and sampling fractions within individual territories covered by the survey, again subject to the limits of available data, and discuss the pattern of variation in the intensity of coverage of the population among territories. The question of the determination of collective farmer sample sizes for individual oblasti is considered more fully in Section 6.

The geographical distribution of the sample within individual territories, rather than among territories, is the subject of Section 7. The main conclusions are summed up in Section 8.

2 The number of territories covered by the FBS

Several Soviet authors state that "some" administrative areas or oblasti are not covered at all by the survey (for example, Levin 1974 p.196, Arkhipova and Bayev 1977, Kuznetsova 1978 p.140). Stolmov (1972 p.65) states that "many" oblasti are excluded.

Matyukha (1973 p.137) reports that:

> almost 100 statistical administrations of oblasti, krai, ASSR's and Republics at present collect family budgets.

Collating this with the report in Soveshchanie... (1969) that

> by the end of 1968, 98 statistical administrations had expanded their samples,

it follows that either 98 or 99 territories were covered by the survey in the early 1970s, so that 55 or 56 territories were excluded - over one-third of the total number. We shall use 99 and 55 as the most likely figures.

The survey is organised by separate Departments of statistical administrations for "families of workers and employees" and for "families of collective farmers". Far from all of those administrations which are

involved in the survey have both types of Budget Survey Department. Maslov, Romanovskaya and Simchera (1978) inform us that workers and/or employees are surveyed in 83 and collective farmers in 56 oblasti - little more than one-half and one-third respectively of all territories. It follows that both workers and/or employees and collective farmers are surveyed together in only 40 territories - just over one-quarter of them (see Table B3.1).

We have information on number of territories covered by individual Republics only for the RSFSR, Belorussia and Kazakhstan.

A meeting of TsSU RSFSR on the budget survey in 1964 (Soveshchanie... 1964) was attended by representatives of 44 oblasti, krai, ASSR's and local Inspectorates. To derive the number of oblast'-level territories represented, we must subtract the estimated number of Inspectorates. The involvement of local Inspectorates in budget survey work is a rare departure from standard practice: there is a speaker from the town of Noginsk, Moscow oblast', but other speakers all come from oblast' offices. The Noginsk Inspectorate has, it seems, been placed experimentally under the direct supervision of TsSU RSFSR, bypassing the Moscow oblast' administration, in order to develop it as a model of exemplary work. The only other local Inspectorate which we know to collect budgets is that of Pechora raion, Komi ASSR (Filippov 1961). It follows that about 42 territories out of 71 in the RSFSR are covered by the survey, a slightly smaller proportion than for the USSR as a whole[2].

The conduct of the survey in five of the six oblasti of Belorussia is alluded to in Soveshchanie... (1969); by implication the sixth oblast' (Grodno) is not covered.

Kazakhstan had in 1958 - as in 1939 (Postnikov 1966) - a sample of 721 collective farmer families (Uchebnoe... 1958 p.215). We are also told in Uchebnoe... (1958) that 80-140 families are selected in each oblast'[3]. By division it follows that 6-8 of the 19 oblasti of Kazakhstan are covered

TABLE B3.1

THE TOTAL NUMBER OF TERRITORIES COVERED BY THE FBS

Number of territories in which workers and/or employees are:	Number of territories in which collective farmers are: Covered	Not covered	Total
Covered	40	43	83
Not covered	16	55	71
Total	56	98	154

by the collective farmers' survey.

3 The identity of the territories covered by the sample

Information on the territorial coverage of the survey is brought together in no single source, and a very incomplete picture has to be pieced together from a wide variety of sources: statistical handbooks[4], reports of meetings connected with the survey (whether of survey staff or of participants), economic research monographs using survey data, comments on the survey by critical economists, speeches at conferences of statisticians, articles on the survey in the general press[5] etc.

In this way 57 of the 99 territories covered by the survey have been identified by name; Table B3.2 gives a regional breakdown, Table B3.3 lists these territories and gives the sources used, and Map 1 shows their location. A further ten territories have been identified as possible survey areas: three where the evidence is conflicting and seven where the pre-war conduct of the survey makes its post-war conduct likely (yellow on the map)[6]. In addition, seven territories have been identified (one with some doubt) for which there is evidence that the survey is <u>not</u> conducted (blue on the map). Thus we have <u>some</u> evidence, one way or the other, regarding 74 territories; regarding the other 80 we know nothing.

For only one-half (37) of the territories regarding which we have any information can we indicate whether both workers/employees and collective farmers are surveyed (24), workers/employees alone (2), collective farmers alone (4) or neither (7). Territorial coverage for workers, employees and collective farmers separately is set out, where known, in Table B3.4 and illustrated by Maps 2 (a), (b) and (c).

4 The overall range of geographical coverage of the sample

The only official reference to a systematic criterion for selecting

TABLE B3.2

NUMBERS OF IDENTIFIED TERRITORIES[1] BY REGION IN WHICH FBS IS CONDUCTED

Region	Total number of territories in region	Number of identified territories in which FBS is conducted		
		Post-war evidence[2]	Pre-war evidence[3]	Total
	(A)	(B)	(C)	(D) = (B)+(C)
RSFSR North-West	8	3 + 1?[4]	0	3 + 1?
Centre	12	8	1	9
Volga-Vyatka	5	2	1	3
Black-Earth Centre	5	2	0	2
Volga	9	4	1	5
North Caucasus	7	2	1	3
Urals	6	3	0	3
West Siberia	6	2 + 1?	1	3 + 1?
East Siberia	5	2	0	2
Far East	7	1 + 1?	0	1 + 1?
Kaliningrad	1	1		1
RSFSR	71	30 + 3?	5	35 + 3?
Ukraine Donets-Dnepr	8	5	0	5
South-West	13	3	2	5
South	4	3	0	3
Ukraine	25	11	2	13
Belorussia	6	5	0	5
The Baltic[5]	3	3	0	3
Moldavia	1	1		1
Georgia[6]	3	1	0	1
Azerbaijan	2	0	0	0
Armenia	1	1		1
Transcaucasia	6	2	0	2

TABLE B3.2 (cont'd)

Region	Total number of territories in region	Number of identified territories in which FBS is conducted		
		Post-war evidence	Pre-war evidence	Total
	(A)	(B)	(C)	(D) = (B)+(C)
Kazakhstan(7)	19	2	0	2
Uzbekistan	12	0	0	0
Kirgizia	3	0	0	0
Tadzhikistan(8)	3	3		3
Turkmenistan	5	0	0	0
Central Asia	42	5	0	5
USSR	154	57 + 3?	7	64 + 3?

Notes:

(1) "Territories" are the lowest level administrative areas in which the FBS is normally organised. In 1978 there were 154 such areas:

(a) Republics subdivided not into oblast'-level units but into units below oblast' level (raiony etc) - the Baltic Republics, Moldavia, Armenia. Here the FBS is conducted by the Republican TsSU;

(b) Oblast'-level units of the other Republics. Here the FBS is conducted by the statistical administrations of these units.

"Oblast'-level units" are units immediately below Republican level in the Republics referred to under (b): oblasti, Krai, Autonomous Soviet Socialist Republics (ASSR's) and areas directly under Republican subordination.

Exceptional cases in which conduct of the FBS is delegated to offices at town or raion level are left out of account.

The Table shows the numbers of territories in which the FBS is conducted and which we are in a position to identify by name. For example, the FBS is known to be conducted in Azerbaijan, Uzbekistan, Kirgizia and Turkmenistan, but as we do not know in which territories of these Republics it is conducted, the Table indicates zero.

(2) Territories for which evidence of the conduct of the FBS exists relating to any time since 1950

TABLE B3.2 (cont'd)

Notes (cont'd):

(3) Territories for which post-war evidence of the conduct of the FBS does not exist, but which are identified by Postnikov (1966) as territories in which the survey of collective farmers was conducted in 1939. It is likely that many of these territories have been covered by the FBS since the war, although direct evidence of this is lacking.

(4) This notation signifies that there is clear evidence for the conduct of the FBS in three territories and contradictory evidence relating to one other territory.

(5) Latvia, Estonia and Lithuania

(6) The identified territory covered by the FBS in Georgia is the area under Republican subordination which includes Tbilisi.

(7) Between six and eight territories are covered by the FBS in Kazakhstan, but the only two which can be identified are East Kazakhstan and Karaganda oblasti.

(8) The "basic agricultural areas", but not the mountainous parts, of each of the three territories of Tadzhikistan - Leninabadskaya and Kulyabskaya oblasti and the area under Republican subordination - are covered by the FBS.

TABLE B3.3

TERRITORIES COVERED BY THE FBS

Code (1)	Territory (by region)	Coverage category (1)	References
	RSFSR North-West		
1	Arkhangelsk	Green	Shvyrkov and Shvyrkova (1973 p 50) for 1952-63
2	Vologda	Green	NKh Vologodskoi oblasti za gody sovetskoi vlasti (1967); NKh RSFSR za 60 let (1977)(2)
3	Leningrad	Green	Iz opyta... (1955); Soveshchanie... (1969); NKh RSFSR za 60 let (1977); statistical handbooks for Leningrad for 1964, 1967 and 1974
4	Murmansk	Blue	Milner and Gilinskaya (1981)
5	Novgorod	White	
6	Pskov	White	
7	Karelian ASSR	White	
8	Komi ASSR	Green?(3)	Filippov (1961); Milner and Gilinskaya (1981)
	RSFSR Centre		
9	Bryansk	White	
10	Vladimir	Green	Vladykin (1958)
11	Ivanovo	Green	Vladykin (1958); Gorod Ivanovo za 100 let (1971); NKh RSFSR za 60 let (1977)(4)
12	Kalinin	Green	Soveshchanie... (1969)
13	Kaluga	White	
14	Kostroma	White	
15	Moscow	Green	Vladykin (1958); Sobranie... (1963); Rovinskaya (1965); Vasilyeva (1965); Soveshchanie... (1969); Sukhova (1969); Shvyrkov and Shvyrkova (1973); Panina (1974); Kotlyarenko (1977); Nemchinova (1977); many statistical handbooks for Moscow

TABLE B3.3 (cont'd)

Code	Territory (by region)	Coverage category	References
16	Orel	Green	NKh SSSR 1922-72, M 1972, pp 383-4
17	Ryazan'	Green	" "
18	Smolensk	Yellow	
19	Tula	Green	Parfenova (1964)
20	Yaroslavl'	Green	Vladykin (1958)
	RSFSR Volga-Vyatka		
21	Gorkii	Green	Krasnoshchekov (1964); Rabochii... (1969 p 231)
22	Kirov	Green	NKh Kirovskoi oblasti (1960); NKh RSFSR za 60 let (1977)
23	Mariisk ASSR	White	
24	Mordov ASSR	Yellow	
25	Chuvash ASSR	White	
	RSFSR Black-Earth Centre		
26	Belgorod	White	
27	Voronezh	Green	Soveshchanie... (1969); NKh RSFSR za 60 let (1977)
28	Kursk	White	
29	Lipetsk	White	
30	Tambov	Green	NKh SSSR 1922-72, M 1972, pp 383-4
	RSFSR Volga		
31	Astrakhan'	White	
32	Volgograd	Green	Vladykin (1958); Krasnoshchekov (1964); Rabochii... (1969 p 231)
33	Kuybyshev	Green	Vladykin (1955); Klyuchenskii (1955); Rabochii... (1969 p 231)
34	Penza	White	
35	Saratov	Green	Rabochii... (1969 p 231)
36	Ulyanovsk	White	
37	Bashkir ASSR	Yellow	
38	Kalmyk ASSR	White	

TABLE B3.3 (cont'd)

Code	Territory (by region)	Coverage category	References
39	Tatar ASSR	Green	Shcherbakov (1951); Vladykin (1958); Soveshchanie... (1964); Parfenova (1964); Karapetyan (1980 p 31); statistical handbooks for the Tatar ASSR for 1960, 1966, 1967, 1970 and 1972
	RSFSR North Caucasus		
40	Krasnodar krai	Green	Soveshchanie... (1964); Sidlyarenko (1964); Khoroshun (1965)
41	Stavropol'	White	
42	Rostov	Green	Krasnoshchekov (1964); Parfenova (1964); Soveshchanie... (1969); Shvyrkov and Shvyrkova (1973 p 50) for 1952-63; Levkova (1976) (5)
43	Dagestan ASSR	White	
44	Kabardino-Balkar ASSR	Yellow	
45	Severo-Osetin ASSR	White	
46	Checheno-Ingush ASSR	White	
	RSFSR Urals		
47	Kurgan	White	
48	Orenburg	Green	Vladykin (1958)
49	Perm'	Green	Vsesoyuznoe soveshchanie... (1954); Lukasheva (1964); Neshitoi and Shatalova (1980)
50	Sverdlovsk	Green	Soveshchanie... (1964); Krasnoshchekov (1964)
51	Chelyabinsk	White	
52	Udmurt ASSR	White	
	RSFSR West Siberia		
53	Altai krai	Yellow	
54	Kemerovo	Green	Maslov (1967)
55	Novosibirsk	Green	Vsesoyuznoe soveshchanie... (1954); Rabochii... (1969 p 231)
56	Omsk	White	

TABLE B3.3 (cont'd)

Code	Territory (by region)	Coverage category	References
57	Tomsk	White	
58	Tyumen'	Green?(6)	Soveshchanie... (1969); Karapetyan (1980 p 31); Milner and Gilinskaya (1981)
	RSFSR East Siberia		
59	Krasnoyarsk	White	
60	Irkutsk	Green	Krasnoshchekov (1964); Kotlyarenko (1977); Nemchinova (1977)
61	Chita	White	
62	Buryat ASSR	Green	Soveshchanie... (1969)
63	Tuvinsk ASSR	White	
	RSFSR Far East		
64	Primorsk krai	Green	Territorialnye... (1966); Kotlyarenko (1977); Nemchinova (1977)
65	Khabarovsk	Blue	Territorialnye... (1966); Milner and Gilinskaya (1981)
66	Amur	Blue	Territorialnye... (1966)
67	Kamchatka	Blue	Territorialnye... (1966); Milner and Gilinskaya (1981)
68	Magadan	Blue	Territorialnye... (1966); Milner and Gilinskaya (1981)
69	Sakhalin	Blue?(7)	Territorialnye... (1966); Karapetyan (1980 p 31)
70	Yakut ASSR	Blue	Territorialnye... (1966); Milner and Gilinskaya (1981)
71	Kaliningrad	Green	Krasnoshchekov (1964)
	Ukraine Donets-Dnepr		
U3	Dnepropetrovsk	Green	Zavrin (1965)
U4	Donetsk	Green	Brainin (1964); Khristovoi (1967); Babayev (1972); Golub (1976)
U7	Zaporozhye	Green	Zavrin (1965)
U10	Kirovograd	White	

TABLE B3.3 (cont'd)

Code	Territory (by region)	Coverage category	References
U12	Lugansk[8]	Green	Kotlyarenko (1977); Nemchinova (1977)
U16	Poltava	White	
U18	Sumii	White	
U20	Kharkov	Green	Podgornov (1957); Zavrin (1965); Zuyeva (1965); Pivovarov (1967); Babayev (1972); Kuzmenkova (1973)
	Ukraine South-West		
U1	Vinnitsa	Yellow	
U2	Volynsk	Green	Solop (1970); Babayev (1972)
U5	Zhitomir	White	
U6	Zakarpatsk	White	
U8	Ivano-Frankov	White	
U9	Kiev	Green	Shkrebel' (1957); Goldman (1963)
U13	Lvov	Green	Krasnoshchekov (1962); Babayev (1967, 1972)
U17	Rovno	White	
U19	Ternopol'	White	
U22	Khmelnitsk	White	
U23	Cherkassii	White	
U24	Chernigov	Yellow	
U25	Chernovtsii	White	
	Ukraine South		
U11	Krym	Green	Soveshchanie... (1969)
U14	Nikolayev	Green	Aleksandrova (1967)
U15	Odessa	Green	Krasnoshchekov (1962); Klebanov (1968)
U21	Kherson	White	
	Belorussia		
B1	Brest	Green	Soveshchanie... (1969); Shamanskii and Zuyev (1976) for 1961-73
B2	Vitebsk	Green	Soveshchanie... (1969)
B3	Gomel	Green	Soveshchanie... (1969); Klimenok (1974)

TABLE B3.3 (cont'd)

Code	Territory (by region)	Coverage category	References
B4	Grodno	Blue[9]	Soveshchanie... (1969)
B5	Minsk	Green	Soveshchanie... (1969); Dedyulya (1970, 1972); Kotlyarenko (1977); Nemchinova (1977)
B6	Mogilev	Green	Vsesoyuznoe... (1959 p 71); Soveshchanie... (1969)
	The Baltic		
	Lithuania	Green	Sukhoruchkina (1970)
	Latvia	Green	Berzkaln (1960); Krasnoshchekov (1962); Rimashevskaya (1965); statistical handbooks for Latvia for 1960 and 1962
	Estonia	Green	Dostizheniya Sovetskoi Estonii za 20 let, Tallin 1960; Krasnoshchekov (1962); Rimashevskaya (1965)
	Moldavia	Green	Statistical handbooks for Moldavia for 1961, 1962 and 1964; Soveshchanie... (1969); Doroganich (1974)
	Transcaucasia		
	Georgia	Green	Morozov (1961); NKh Gruzinskoi SSR (1963)
	Azerbaijan	Green	Iz opyta... (1955); Dostizheniya Sovetskogo Azerbaidzhana za 40 let v tsifrakh, Baku 1960; Razvitie NKh AzSSR i rost materialnogo i kulturnogo urovnya zhizni naroda, Baku 1961; Karapetyan (1980 p 31)
	Armenia	Green	Krasnoshchekov (1962); Rimashevskaya (1965); Kotlyarenko (1977)

TABLE B3.3 (cont'd)

Territory (by region)	Coverage category	References
Central Asia		
Kazakhstan(10)		
Kazakhstan as a whole	Striped yellow	Krasnoshchekov (1962); Soveshchanie... (1970)
East Kazakhstan	Green	Konovalov (1965)
Karaganda	Green	Rabochii... (1969 p 231)
Uzbekistan	Green	Krasnoshchekov (1962); Soveshchanie... (1970); Shvyrkov and Shvyrkova (1973 p 66)
Kirgizia	Green	Kalinichenko (1963); Altunina (1974); Kirgizstan za 50 let (1974); NKh Kirgizskoi SSR (1975, 1976, 1977); Shiriyazdanov (1979 p 372)
Tadzhikistan(11)	Green	Krasnoshchekov (1962); NKh Tadzhikskoi SSR (1961, 1962)
Turkmenistan	Green	Krasnoshchekov (1962)

Notes:

(1) The codes are the territory codes used in Maps 1 and 2 (a), (b) and (c). The coverage categories are those used in Map 1.

(2) NKh - Narodnoe Khozyaistvo

(3) Coverage is reported by Filippov (1961), but Milner and Gilinskaya (1981) indicate no coverage.

(4) Reference to the town of Furmanov in Ivanovo oblast'.

(5) Reference to the town of Taganrog in Rostov oblast'

(6) Soveshchanie... (1969) and Karapetyan (1980) indicate coverage, but Milner and Gilinskaya (1981) indicate no coverage.

(7) Karapetyan (1980) implies coverage; Territorialnye... (1966) indicates no coverage.

(8) Lugansk was formerly Voroshilovgrad.

(9) Non-coverage implied by context in Soveshchanie... (1969)

(10) Uchebnoe... (1958 p 214) gives oblast' sample sizes for Kazakhstan which imply that collective farmers are covered in

TABLE B3.3 (cont'd)

Notes (cont'd):

between six and eight *oblasti*.

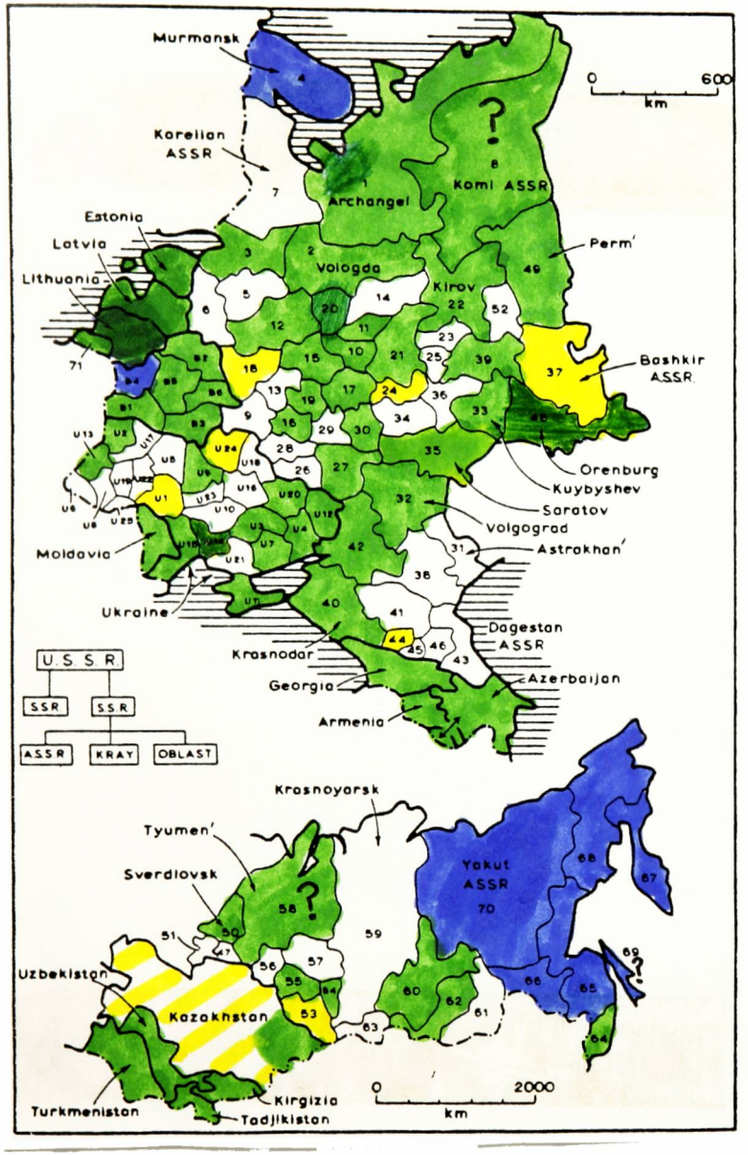
(11) *NKh Tadzhikskoi SSR* (1961) states that the collective farms covered in Tadzhikistan are evenly distributed through the "basic agricultural regions" of the Republic - that is, in Leninabadskaya *oblast'*, in the Kulyabskaya group of *raiony* of Kulyabskaya *oblast'*, and in the Gissarskaya and Vakhshskaya valleys (in the area under Republican subordination). The more mountainous areas of the Republic are not covered (Postnikov 1961).

MAP 1

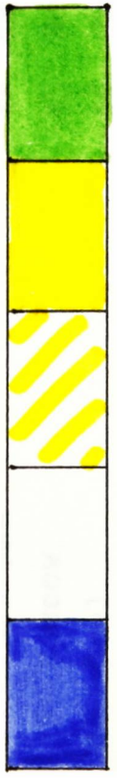
GEOGRAPHICAL COVERAGE OF THE FAMILY BUDGET SURVEY

Politico-administrative areas of the USSR (small Republics; oblasti, krai and ASSR's of large Republics) have been divided into the following five categories on the basis of very incomplete information from disparate sources.

We consider it likely that the survey is conducted in all the "green" and "yellow" areas, and in about one-third of the "white" areas (we do not know which). In some areas, however, coverage applies only to part of the period since 1952 and/or is on a very small scale (eg Komi ASSR).



Source of outline map: J P Cole, A geography of the USSR (Butterworth 1970), Fig. 2.3, p 21



Indication(s) that survey has been conducted here in the period since 1952.

As survey was conducted here in 1939*, its post-War reconstruction is likely, although direct evidence is lacking.

*Postnikov 1966

"Yellow" situation in Kazakhstan, but probably only some of the oblasti are covered.

(Indication that survey is conducted in East Kazakhstan oblast')

No information available on survey in these areas.

Indication that survey is NOT conducted in these areas**.

**Territorialnye problemy... 1966, Kirber and Sitinskaya 1981

?

References contradictory: one or two support the colour shown, but one points in the opposite direction.

KEY TO TERRITORY CODES IN MAPS 1 AND 2 (a), (b), (c)

RSFSR North-West

1 Arkhangelsk
2 Vologda
3 Leningrad
4 Murmansk
5 Novgorod
6 Pskov
7 Karelian ASSR
8 Komi ASSR

RSFSR Centre

9 Bryansk
10 Vladimir
11 Ivanovo
12 Kalinin
13 Kaluga
14 Kostroma
15 Moscow
16 Orel
17 Ryazan'
18 Smolensk
19 Tula
20 Yaroslavl'

RSFSR Volga-Vyatka

21 Gorkii
22 Kirov
23 Mariisk ASSR
24 Mordov ASSR
25 Chuvash ASSR

RSFSR Black-Earth Centre

26 Belgorod
27 Voronezh
28 Kursk
29 Lipetsk
30 Tambov

RSFSR Volga

31 Astrakhan'
32 Volgograd
33 Kuybyshev
34 Penza
35 Saratov
36 Ulyanovsk
37 Bashkir ASSR
38 Kalmyk ASSR
39 Tatar ASSR

RSFSR North Caucasus

40 Krasnodar krai
41 Stavropol'
42 Rostov
43 Dagestan ASSR
44 Kabardino-Balkar ASSR
45 Severo-Osetin ASSR
46 Checheno-Ingush ASSR

RSFSR Urals

47 Kurgan
48 Orenburg
49 Perm'
50 Sverdlovsk
51 Chelyabinsk
52 Udmurt ASSR

RSFSR West Siberia

53 Altai krai
54 Kemerovo
55 Novosibirsk
56 Omsk
57 Tomsk
58 Tyumen'

KEY TO TERRITORY CODES (cont'd)

RSFSR East Siberia

59 Krasnoyarsk
60 Irkutsk
61 Chita
62 Buryat ASSR
63 Tuvinsk ASSR

RSFSR Far East

64 Primorsk krai
65 Khabarovsk
66 Amur
67 Kamchatka
68 Magadan
69 Sakhalin
70 Yakut ASSR

71 Kaliningrad

UKRAINE

U1 Vinnitsa
U2 Volynsk
U3 Dnepropetrovsk
U4 Donetsk
U5 Zhitomir
U6 Zakarpatsk
U7 Zaporozhye
U8 Ivano-Frankov
U9 Kiev
U10 Kirovograd
U11 Krym
U12 Lugansk
U13 Lvov
U14 Nikolayev
U15 Odessa
U16 Poltava
U17 Rovno
U18 Sumii
U19 Ternopol'
U20 Kharkov
U21 Kherson
U22 Khmelnitsk
U23 Cherkassii
U24 Chernigov
U25 Chernovtsii

BELORUSSIA

B1 Brest
B2 Vitebsk
B3 Gomel
B4 Grodno
B5 Minsk
B6 Mogilev

Source: Cole (1970), Tables 2.2 and 2.3, pp 22-4

TABLE B3.4

SOCIAL GROUPS COVERED BY FBS IN "GREEN" POLITICO-ADMINISTRATIVE TERRITORIES

Code: * evidence that the social group is covered
x evidence that the social group is not covered
- no evidence as to whether or not the social group is covered
? sources contradictory regarding coverage of the social group (most plausible assessment indicated)

Territory	Workers	Employees	Collective farmers
RSFSR North-West			
Arkhangelsk	*	-	-
Vologda	-	-	*
Leningrad	*	*	*
Komi ASSR	x	x	*?
RSFSR Centre			
Vladimir	*	-	-
Ivanovo	*	*	-
Kalinin	-	-	-
Moscow	*	*	*
Orel	-	-	*
Ryazan'	-	-	*
Tula	-	*	-
Yaroslavl'	-	-	-
RSFSR Volga-Vyatka			
Gorkii	*	-	-
Kirov	*	-	*
RSFSR Black-Earth Centre			
Voronezh	-	-	*
Tambov	-	-	*
RSFSR Volga			
Volgograd	*	-	-
Kuybyshev	*	*	x
Saratov	*	-	-
Tatar ASSR	*	*	*

TABLE B3.4 (cont'd)

Territory	Workers	Employees	Collective farmers
RSFSR North Caucasus			
Krasnodar krai(1)	*	*	*
Rostov	*	*	-
RSFSR Urals			
Orenburg	-	-	-
Perm'	*	-	*
Sverdlovsk	*	-	-
RSFSR West Siberia			
Kemerovo	*	-	-
Novosibirsk	*	*	*
Tyumen'	*?	x	x
RSFSR East Siberia			
Irkutsk	*	-	-
Buryat ASSR	-	-	-
RSFSR Far East			
Primorsk krai	*	x	-
RSFSR Kaliningrad	-	-	-
Ukraine Donets-Dnepr			
Dnepropetrovsk	-	-	-
Donetsk(1)	*	*	x
Zaporozhye	-	-	-
Lugansk	*	-	-
Kharkov	*	-	*
Ukraine South-West			
Volynsk(1)	x	x	*
Kiev	*	*	*
Lvov(1)	*	*	*
Ukraine South			
Krym	-	-	-
Nikolayev	*	-	-
Odessa(1)	*	*	*

TABLE B3.4 (cont'd)

Territory	Workers	Employees	Collective farmers
Belorussia			
Brest	*	*	*
Vitebsk	x	x	*
Gomel	*	-	*
Minsk(1)	*	*	*
Mogilev	x	x	*
Lithuania	*	*	*
Latvia	*	*	*
Estonia	*	*	*
Moldavia	*	-	*
Georgia	*	-	*
Azerbaijan	*	-	*
Armenia	*	*	*
Kazakhstan			
East Kazakhstan(1)	*	x	*
Karaganda	*	-	-
Uzbekistan	-	-	*
Kirgizia	*	*	*
Tadzhikistan	*	-	*
Turkmenistan	*	-	*

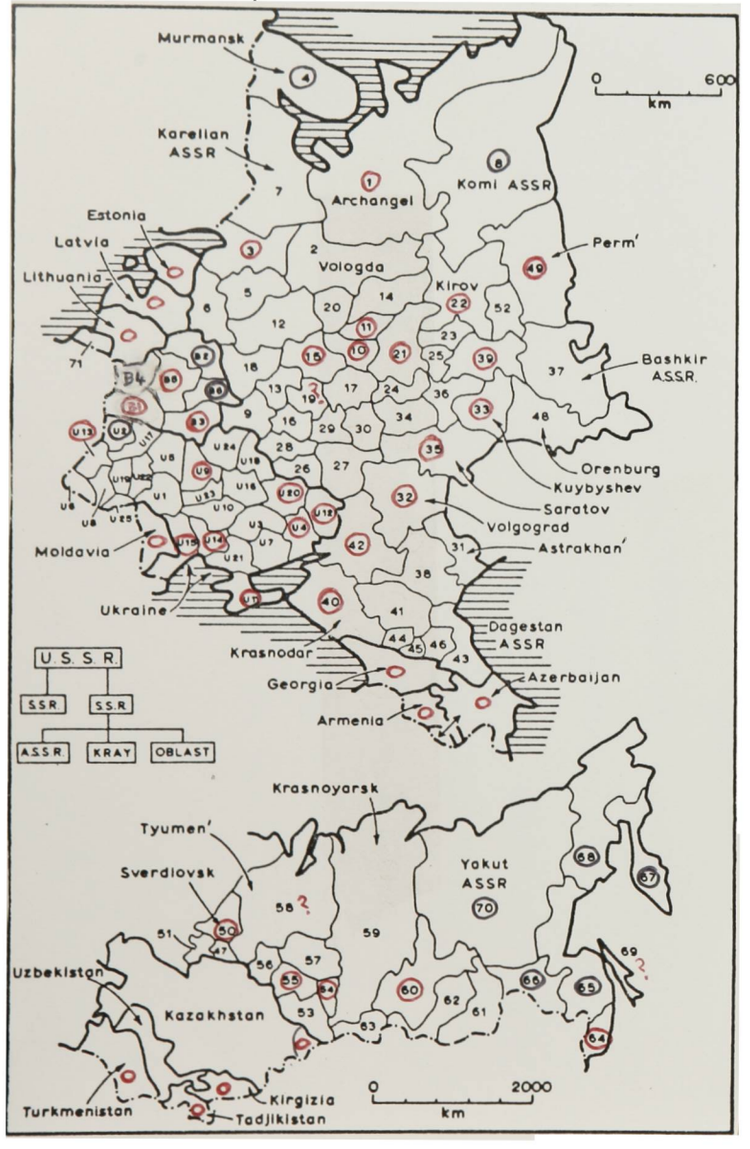
Note: (1) Evidence that workers on State farms are covered

MAPS 2 (a), (b), (c)

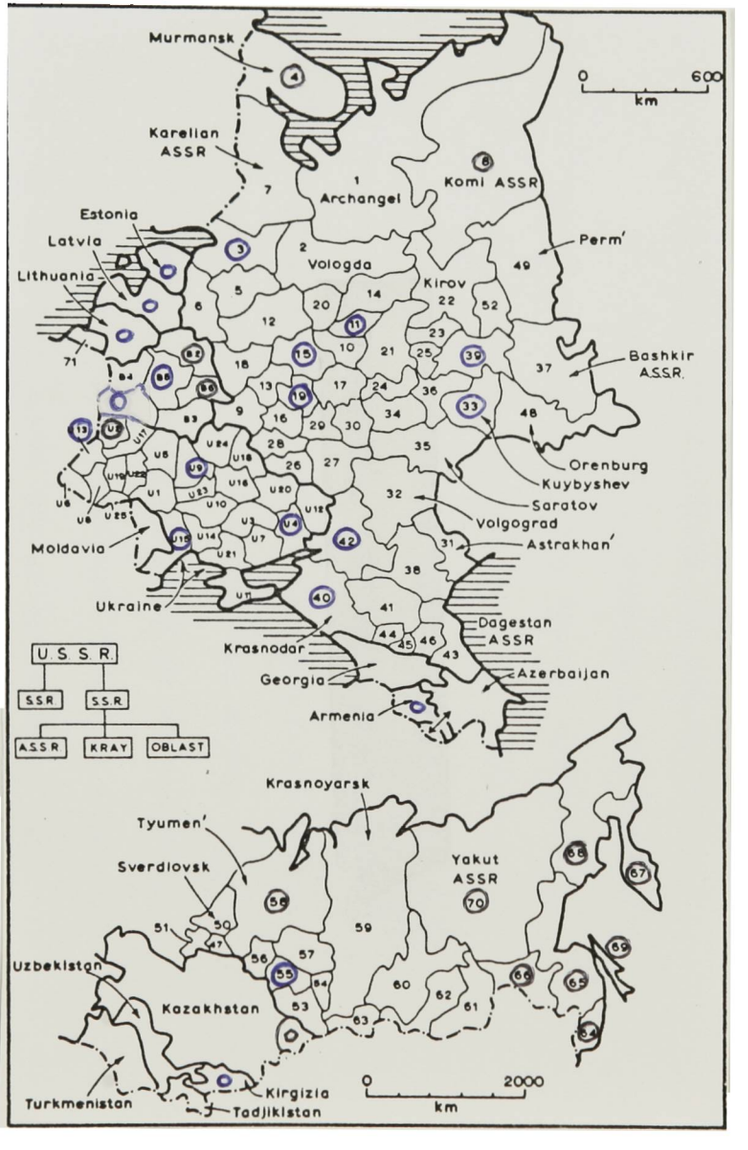
GEOGRAPHICAL COVERAGE OF THE FAMILY BUDGET SURVEY BY SOCIAL GROUP

Coloured ring - evidence that social group covered in the area
? - conflicting evidence: social group may be covered
No marking - no indication whether or not social group covered
Black ring - evidence that social group not covered in the area

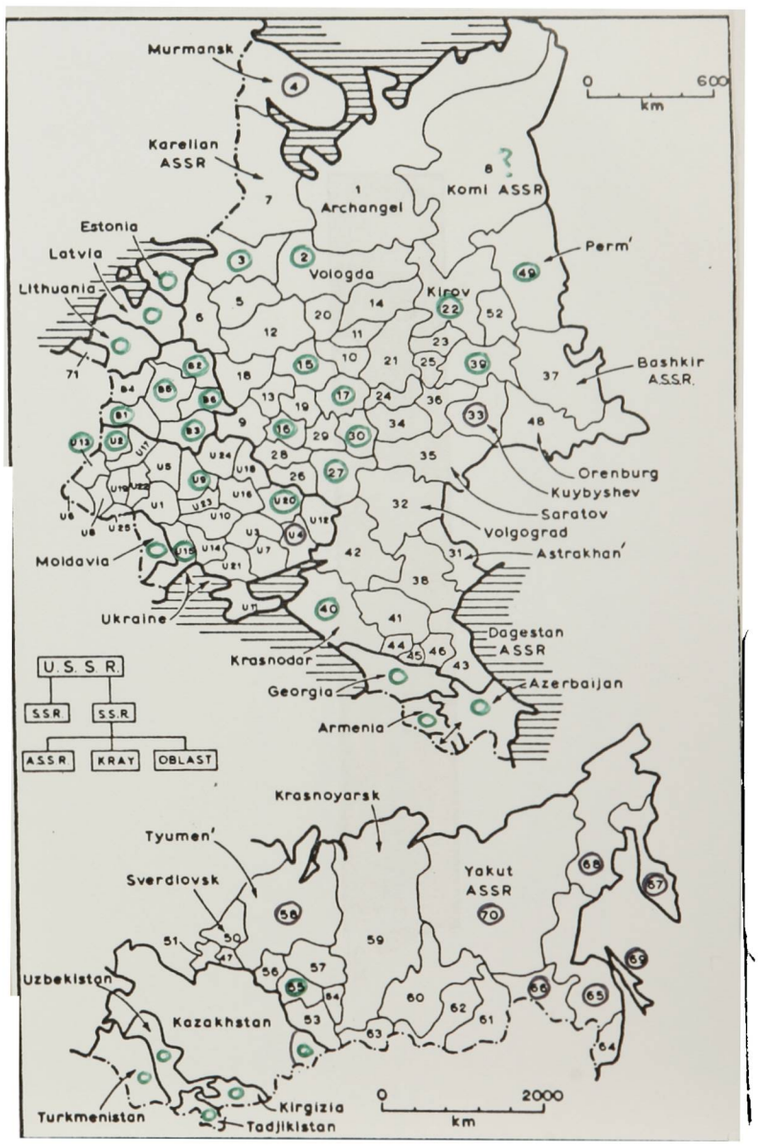
(a) Manual workers



(b) Employees



(c) Collective farmers



territories for inclusion in the FBS is that given in a textbook of agricultural statistics (Statistika... 1973 pp.378-9). It is argued that an oblast' sample of collective farmers must contain at least 5-600 families in order to yield stable indicators (see Section 6). As the collective farmer sample for the USSR is not large enough to cover all oblasti at this minimum level, it is necessary to select oblasti for inclusion. The oblasti selected are those "most typical of each natural-economic zone".

The information on the territorial coverage of the sample presented in Section 3 does not, however, support the claim that the various natural-economic zones of the USSR are more or less evenly covered. The central Russian region is clearly well represented both in the survey of workers and employees and in that of collective farmers, while five out of the six oblasti in Belorussia are included in the survey of collective farmers. All three of the Baltic Republics are covered. On the other hand, the sparsely populated regions characterised by extreme climatic conditions - the Far North and the Far East and much of Kazakhstan and Central Asia - are very poorly covered by the survey.

Thus, in the Far Eastern region only the most accessible area around Vladivostok, Primorskii krai, is covered:

> Is it possible to assess the income and consumption levels of all workers and employees in the Far East from the budgets of 200 families in Primorskii krai?
>
> (Territorialnye... 1966)

Milner and Gilinskaya (1981) state that the survey is mainly restricted to areas of long-established habitation (obzhitye raiony), and does not cover such areas of economic potential as Tyumen' (West Siberia); Magadan, Kamchatka, Yakutia and Khabarovsk (Far East); Murmansk and the Komi ASSR (Far North of European Russia). This list does not seem to be completely accurate: two of our sources suggest that oil workers in Tyumen' are

covered, and there is also a report of the conduct of the FBS on a small scale in Pechora raion of the Komi ASSR (Filippov 1961)[7]. However, there is no doubt that Siberia, the Far North and the Far East are covered much less intensively than more temperate regions.

Similarly, the mountainous parts of Tadzhikistan in Central Asia are not covered by the sample (Postnikov 1961), and it is to be doubted whether desert regions are covered either. In general, the sample tends not to cover those regions where the conduct of the survey would be impeded by dispersed populations, harsh climate and difficult communications.

5 Sample sizes and sampling fractions in the territories covered by the sample

An assessment of the territorial coverage of the survey should of course take into account the intensity of coverage of the population on the territories covered as well as the number and identity of those territories. However, the information we have on the sizes of samples in parts of the USSR is limited to ten of the Union Republics, three cities (Moscow, Leningrad and Riga) and twelve oblast'-level territories. Generalisations drawn on the basis of such meagre data must be regarded as provisional.

The sample size data, and the sources from which they are derived, are set out in Table B3.5[8]. Sampling fractions derived from these sample sizes are set out in Table B3.6, while in Table B3.7 territories are classified into three groups - those which the survey covers more intensively than the USSR as a whole, those covered about as intensively and those covered less intensively.

Sampling fractions have been calculated as follows:

$$(SF) = (SS) \times (CF) \times \frac{(AFS)}{(POP)},$$

TABLE B3.5
SAMPLE SIZES BY TERRITORY

Code: w families of workers (including workers on State farms, except where shown separately)
e families of employees (including ITR)
s families of workers on State farms
k families of collective farmers
t total: all families in the sample

Territory	Sample sizes	References
UNION REPUBLICS:		
RSFSR 1959	10000 w, 13000 k	Shvyrkov and Shvyrkova (1973 pp 56, 68)
1963-4	29000 t	Soveshchanie... (1964), Krasnoshchekov (1964)
	10000 w	Shvyrkov and Shvyrkova (1973)
	2859 e	Parfenova (1964), Sabinin (1966)(1)
1966	3315 e	" "
Belorussia 1958	300 w	Dostizheniya sovetskoi Belorussii za 40 let, Minsk 1958, p 143 (2)
1964	362 we, 1580 k	Drits (1968)
Moldavia 1953	64 t	Doroganich (1974 p 57fn)
1958	103 t (3)	"
1960	103 w, 504 k	NKh MoldSSR v 1960 g, Kishinev 1961, p 274
1961	121 w, 528 k	Moldavskaya SSR v tsifrakh v 1961 godu, Kishinev 1962, p 277
Latvia 1965	100-150 we	Rimashevskaya (1965)
1975	≥245 we	Venetskii and Bauman (1975)
Estonia 1965	100-150 we	Rimashevskaya (1965)
Georgia 1961	0·14% k	Morozov (1961)
Armenia 1965	100-150 we	Rimashevskaya (1965)

TABLE B3.5 (cont'd)

Territory	Sample sizes	References
UNION REPUBLICS (cont'd):		
Kazakhstan		
1958	720 k	Uchebnoe posobie po... , M 1958, p 214
Kirgizia		
1973-6	1019 t	Kirgizstan za 50 let, Frunze 1974, p 243; NKh KirgSSR v 1974 g, Frunze 1975, p 251; NKh KirgSSR v 1975 g, Frunze 1976, p 278; Kirgizstan za gody sovetskoi vlasti, Frunze 1977, p 149
Tadzhikistan		
1960	121 w, 360 k	NKh TadzhSSR v 1960 g, Dushanbe 1961, p 17
1961	134 w, 441 k	NKh TadzhSSR v 1961 g, Dushanbe 1962, p 20
CITIES:		
Leningrad		
1957	839 w, 346 e	NKh goroda Leningrada, Leningrad 1957, p 68 (2)
Moscow 1956	1057 w	Moskva - razvitie khozyaistva i kultury goroda, M 1958, p 68
1958	1077 w	Uchebnoe posobie po... , M 1958, p 215
1959	1200 w	Shvyrkov and Shvyrkova (1973 p 56)
1962	1800 we	Sobranie... (1963)
Riga (Latvia) 1975	245 we	Venetskii and Bauman (1975)
OBLASTI ETC:		
RSFSR		
Moscow 1959	310 k	Shvyrkov and Shvyrkova (1973 p 76)
1968	1200 k	Vsesoyuznoe... (1969 p 395)
Kuybyshev		
1955	255 w, 76 e	Vladykin (1955)
Kemerovo		
1967	~400 w	Maslov (1967 p 155)

TABLE B3.5 (cont'd)

Territory		Sample sizes	References
OBLASTI ETC (cont'd):			
Krasnodar krai	1963	267 we, 75 s, 550 k	Sidlyarenko (1964)
Komi ASSR	1961	50 k	Filippov (1961)
Primorsk krai	1966	200 we	Territorialnye... (1966)
BELORUSSIA			
Minsk	1972	125 w, 25 e, 50 s, 325 k	Dedyulya (1972)
Gomel	1974	525 t	Klimenok (1974)
UKRAINE			
Kiev	1957	506 k	Shkrebel' (1957)
Lvov	1967	660 t	Babayev (1967)
	1972	425 w, 25 s, 500 k	Babayev (1972)
Odessa	1968	278 we(4), 22 s, 500 k	Klebanov (1968)
KAZAKHSTAN			
East Kazakhstan	1965	125 w, 25 s, 50 k	Konovalov (1965)

Notes: (1) For the method of imputation from these sources, see Chapter B4 Section 3.

(2) Cited by Goldman (1963)

(3) Uchebnoe... (1958) gives 105 w.

(4) Klebanov gives 250 we, but this excludes a sample of teachers and medical personnel which we estimate as 28 by subtracting the other groups from the total sample of 800.

TABLE B3.6

SAMPLING FRACTIONS BY TERRITORY

		Budgets collected per 1000 families		
		of urban population	of rural population	of overall population
USSR (for comparison)				
	1952	1·1	0·9	1·0
	1965	0·7	1·0	0·8
	1981	0·6	1·1	0·8
UNION REPUBLICS:				
RSFSR[1]	1959	0·7	0·9	0·8
	1963-4	0·6	1·0	0·8
Far East region of the RSFSR	1966	0·2	0·0	0·1
Belorussia	1958	0·5		
	1964[2]	0·3	1·1	0·8
Moldavia	1953-8			0·1
	1960-1	0·7	0·9 - 1·0	0·8
Latvia	1965	0·3		
	1975	⩾0·6		
Estonia	1965	0·5		
Georgia	1961		1·4	
Armenia	1965	0·5		
Kazakhstan	1958		0·6	
Kirgizia	1973-4			1·5
	1975-6			1·4
Tadzhikistan	1960	1·0	1·5	1·3
	1961	1·0	1·8	1·5
CITIES:				
Leningrad	1957	1·4		
Moscow	1956	1·0		
	1959	0·9		
	1962	1·0		
Riga (Latvia)	1975	1·0		

TABLE B3.6 (cont'd)

		Budgets collected per 1000 families		
		of urban population	of rural population	of overall population
OBLASTI ETC:				
RSFSR Moscow	1959		0·6	
	1968		2·5	
Kuybyshev	1955	0·9		
Kemerovo	1967	0·8		
Krasnodar krai	1963	0·5	1·1	0·8
Komi ASSR(3)	1961		0·6	
Primorsk krai	1966	0·6	0·0	0·4
BELORUSSIA				
Minsk	1972	0·4	1·3	0·7
Gomel	1974			1·2
UKRAINE				
Kiev	1957		1·5	
Lvov	1967			1·0
	1972	1·2	1·5	1·3
Odessa	1968	0·9	1·9	1·2
KAZAKHSTAN				
East Kazakhstan	1965	1·0	1·0	1·0

Notes: (1) Shvyrkov (1965) gives sampling fractions for the RSFSR in 1960 of 0·15 per cent for workers' families and 0·18 per cent for collective farmer families - that is, after taking employees into account, about 2 per 1000. Such high figures cannot be reconciled with our calculations. Shvyrkov is possibly giving sampling fractions with respect to territories and branches covered by the survey only.

(2) Based on sources for 1958 and 1968

(3) Assuming that the survey covers only Pechora raion

TABLE B3.7

CLASSIFICATION OF TERRITORIES BY RANGE OF SAMPLING FRACTION

	Territories for which sampling fractions are:		
	greater than those for the USSR as a whole	about the same as those for the USSR as a whole	smaller than those for the USSR as a whole
Union Republics	Rural Georgia Tadzhikistan Kirgizia	RSFSR Rural Belorussia Moldavia	Urban Estonia Urban Latvia Urban Armenia Urban Belorussia Rural Kazakhstan Far East region of the RSFSR
Cities	Leningrad Moscow Riga (1975)		
Oblasti in the RSFSR	Rural Moscow oblast' (1968)	Kuybyshev Kemerovo Krasnodar krai	Rural Moscow oblast' (1953-8) Rural Komi ASSR Primorsk krai
Oblasti in Belorussia	Gomel	Minsk	
Oblasti in the Ukraine	Rural Kiev Lvov Odessa Rural Nikolayev* Rural Dnepropetrovsk* Rural Kharkov* Rural Chernigov*	Rural Vinnitsa*	
Oblast' in Kazakhstan		East Kazakhstan	

* classified on the basis of 1939 sample sizes (Postnikov 1966 p 53), probably a fairly reliable indicator of post-war sample sizes.

where

(SF) is the sampling fraction for a territory,

(SS) is the sample size given for the territory in Table B3.5,

(CF) is a conversion factor given in Table B3.9,

(AFS) is an average family size given in Table B3.8, and

(POP) is the population of the territory as given by the appropriate statistical handbook.

As sample sizes relate usually to "workers" and "collective farmers", while population figures relate to urban and rural parts of territories, it is necessary to apply conversion factors to sample sizes to obtain estimates of urban and rural sample sizes comparable with population data. To obtain the urban sample from the workers' sample, we must add in urban employees and subtract rural workers (mainly workers on State farms). The rural workers must be added to the collective farmers to give the rural sample. The sample breakdown necessary for the calculation of these factors is available only for the USSR as a whole, so that we have to ignore the great geographical variation in such sample ratios as (workers:employees) and (collective farmers: workers on State farms). This is the main source of imprecision in the calculated sampling fractions, and why it would be spurious precision to present them to a greater number of significant digits.

For comparability with samples measured in terms of families, population data must be converted into numbers of families. We have used average urban and rural family sizes by Republic derived from the population censuses of 1959 and 1970. Interpolations and extrapolations in time are not here a serious source of error in view of the slow rates of change. It is conceivable, however, that substantial error has resulted from not being able to use data differentiated by oblast' etc.

Turning to the sampling fractions presented in Table B3.6, the wide range over which they are spread - from 0.1 per 1,000 families in Moldavia

TABLE B3.8
AVERAGE FAMILY SIZES BY UNION REPUBLIC

Union Republic	Average family size		Average urban family size		Average rural family size	
	1959	1970	1959	1970	1959	1970
RSFSR	3·6	3·5	3·5	3·4	3·8	3·8
Ukraine	3·5	3·4	3·4	3·3	3·7	3·6
Belorussia	3·7	3·6	3·5	3·5	3·8	3·7
Lithuania	3·6	3·4	3·4	3·4	3·7	3·5
Latvia	3·2	3·2	3·1	3·2	3·2	3·2
Estonia	3·1	3·1	3·1	3·2	3·1	3·1
Moldavia	3·8	3·8	3·5	3·4	3·9	3·9
Armenia	4·8	5·0	4·5	4·7	5·1	5·5
Kazakhstan	4·1	4·3	3·9	3·9	4·3	4·8
Kirgizia	4·2	4·6	3·9	4·0	4·4	5·1
Tadzhikistan	4·7	5·4	4·1	4·5	5·1	6·0
USSR	3·7	3·7	3·5	3·5	3·9	4·0

Source: Roganova (1976, p 265, Table 4), taken from the population censuses of 1959 and 1970. Figures for Georgia, Azerbaijan and Turkmenistan omitted, as not used in the calculations of sampling fractions.

TABLE B3.9

CONVERSION FACTORS USED FOR THE DERIVATION OF URBAN AND RURAL SAMPLE SIZES

Factors used for the derivation of urban sample sizes from the sizes of manual worker samples

- 1954	1·1
1955 - 1959	1·2
1960 - 1963	1·3
1964 - 1975	1·4
1976 - 1979	1·3
1980 - 1981	1·2

Factors used for the derivation of rural sample sizes from the sizes of collective farmer samples

- 1962	1·05
1963 - 1968	1·10
1969 - 1975	1·15
1976 - 1979	1·20
1980 - 1981	1·25

1953-8 to 2.5 in rural Moscow oblast' 1968[9] is a striking indication of the unevenness of survey coverage.

Out of twelve territories for which we are able to compare urban with rural sampling fractions, the rural sampling fraction is greater in eleven (and in the twelfth they are about equal), suggesting that the denser coverage of rural areas is a consistent feature of the survey throughout the USSR.

The classification of territories given in Table B3.7 is broadly consistent with the complaints by Rimashevskaya (1965) and others that peripheral areas of the USSR are poorly covered by comparison with the central areas. Certainly Moscow and Leningrad have more than their "fair share" of the sample, the Baltic Republics, Armenia, Kazakhstan and the Komi ASSR less than theirs. However, there are quite a few deviations from this simple pattern. For example, it seems that the rural part of Moscow oblast' was not well covered in 1959 (though it was in 1968), while the Trans-caucasian Republic of Georgia and the Central Asian Republics of Tadzhikistan and Kirgizia in fact have relatively large sampling fractions.

The huge variation in sampling fraction among different peripheral Republics is confirmed by the statement of Bredov and Levin (1971) that sample coverage per million inhabitants in Estonia is four times greater than in Kazakhstan, and in Turkmenia is 2.7 times greater than in Uzbekistan[10].

6 The determination of oblast' sample sizes in the survey of collective farmers

Postnikov (1953), Matyukha (1966) and Karapetyan, Rimashevskaya and Shlyapentokh (1967) all mention the practice of setting samples of collective farmer families at approximately the same size in different oblasti. Matyukha states that this standard sample size is 500 families while Postnikov reports 506 "in most krai and oblasti ". This is consistent with

the following evidence:

(a) For 1939 we have a list of collective farmer sample sizes in 26 oblasti of the RSFSR and seven of the Ukraine (Postnikov 1966). Of the 33 samples, no less than 22 consist of exactly 504 families, including all the Ukrainian samples. Six more are approximately of size 500 (two of size 525, four of size 576). In the three territories for which we know both 1939 and post-war sample sizes - Odessa oblast' (Klebanov 1968), Kiev oblast' (Shkrebel' 1957) and Kazakhstan (sample of 720 both in 1939 and in 1958 according to Uchebnoe posobie... 1958) - there is virtually no difference between the two. This is consistent with the hypothesis that the system of standard sample size has persisted.

(b) We have references to post-war sample sizes of about 500 in Odessa and Kiev oblasti (see Table B3.10). The survey covers twenty collective farms in Lvov oblast' (Babayev 1972), implying a sample size again of about 500. Shvyrkov (1965) gives data relating to an unidentified "Southern oblast'" in 1960 from which one can infer a sample size of 504[11]. We also know that Moldavia - a Republic of oblast' size - had a sample of about 500 in 1960-1.

(c) The common use of a standard sample size of about 500 is consistent with many of the examples used by statistics textbooks to explain the method of sample selection. Thus, four out of seven of the textbooks listed in Table B3.11 cite a sample size of about 500. These illustrations, presented as hypothetical, seem in fact to be real and typical cases.

As oblasti vary greatly in the size of their rural populations, use of a standard sample size leads to wide variation in sampling fraction. We illustrate this in respect of some oblasti in the Ukraine for which we have sample size information and the Moldavian Republic, apparently treated in the same way as an oblast'. The sampling fractions derived in Table B3.10 range from 0.9 per 1,000 rural families in Moldavia, with a rural

TABLE B3.10

CONSEQUENCES OF STANDARD COLLECTIVE FARMER SAMPLE SIZES IN UKRAINIAN OBLASTI AND IN MOLDAVIA FOR SAMPLING FRACTIONS

Oblast' or Republic	Rural population[1] in 1959		Sample sizes[2] (year)	Sampling fractions (per 1000 rural families in 1959)
	persons '000	families '000		
Nikolayev	613	166	504 (1939)[3]	3·2
Dnepropetrovsk	798	216	504 (1939)[3]	2·45
Kharkov	904	244	504 (1939)[3]	2·2
Odessa	1064	288	504 (1939)[3] 536 (1968)[4]	1·9
Chernigov	1199	324	504 (1939)[3]	1·6
Kiev	1278	345	504 (1939)[3] 506 (1957)[5]	1·5
Lvov	1270	343	~ 500 (1972)[6]	1·5
Vinnitsa	1774	479	504 (1939)[3]	1·1
Moldavian SSR	2210	567	504 (1960)[7] 528 (1961)[8]	0·9 1·0

Notes

(1) Populations in persons from Narodne Gospodarstvo Ukrainskoi RSR v 1959 rotsi, Kiev 1960 (p 16), converted to numbers of families using the average rural family sizes for the Ukraine (3·7) and Moldavia (3·9) as given by Roganova (1976). Both of these sources derive from the 1959 population census.

(2) We treat sample sizes for various years as proxies for 1959 sample sizes, relating them all to 1959 populations. This is justified in an illustrative exercise in view of the evidence for stability of sample sizes over time (as in Odessa and Kiev oblasti).

In calculating sampling fractions we apply a corrective factor of 1·05 to sample sizes to allow for the presence in rural samples of a few non-collective-farmer families. The factor is based on USSR data (Table B3.9), but agrees approximately with detailed data available for Odessa oblast' from Klebanov (1968).

TABLE B3.10

Notes (cont'd)

(3) Postnikov (1966 p 53)
(4) Klebanov (1968)
(5) Shkrebel' (1957)
(6) Babayev (1972)
(7) Narodnoe Khozyaistvo Moldavskoi SSR v 1960 godu, Kishinev 1961 (p 274)
(8) Moldavskaya SSR v tsifrakh v 1961 godu, Kishinev 1962 (p 277)

TABLE B3.11
TYPICAL SIZES OF OBLAST' SAMPLES OF COLLECTIVE FARMER FAMILIES

Textbook reference	Sample size
Krylov (1957 p 80)	506
Uchebnoe posobie po otdelnym... (1958 p 199)	506
Kurs ekonomicheskoi statistiki (1961 p 392)	1125
Matyukha (1967 p 24)	500
Barbashov et al (1968 p 159)	500
Sergeyev (1968 p 407)	375
Dolgushevskii and Khristich (1976 p 341)	757

Note: The references are to illustrations in statistics textbooks of the method of selection of the FBS collective farmer sample for an oblast'. In each textbook a hypothetical sample size x is given by an expression like: "Let us suppose that we are to select x collective farmer families in an oblast'".

population well in excess of two million, to 3.2 per 1,000 in Nikolayev oblast', with scarcely 600,000 rural inhabitants.

Matyukha (1966), Barbashov et al (1968 p.157) and Statistika...(1973) inform us that, in calculating survey results for Republics and the USSR as a whole, collective farmer data are reweighted in order to correct for the non-proportionality of the distribution of the sample among oblasti. The weights used are the numbers of collective farmer families in different oblasti, taken from the annual census of private livestock. Thus the variation in sampling fraction does not seriously distort global data[12].

The use of a standard sample size is justified on the grounds that reliable data are needed at the oblast' level for various economic calculations (see Chapter D2). In order to yield "sufficiently stable indicators" for these calculations, oblast' samples must be of a certain minimum size, however small the oblast' may be. Different authors cite different minima: Barbashov et al (1968 p.157) consider that 400 families is adequate; Statistika... (1973 p.378) gives "500-600 families"; Postnikov (1953) states that "calculations and practice show that 500-900 budgets ... yield stable indicators".

To the extent that sampling error does depend on the absolute sample size, the approach is a sound one, though the use of a standard sample size may also be attractive by virtue of its administrative simplicity. Karapetyan, Rimashevskaya and Sidlyarenko (1967) point out that sampling error depends also on the variability of income and consumption within oblast' populations: the greater the variability, the larger a sample required for estimates of given precision. They urge that sample sizes be adjusted accordingly. Shvyrkov (1965) takes a similar view: "The practice of observing the same number of collective farmer budgets in each oblast' must be discontinued, for it leads to varying levels of data reliability". He suggests that size of population, the level of its material welfare and the

seasonality of its consumption all be taken into account when determining oblast' sample sizes. It is not clear, however, how substantial the practical gain from such refinements would be.

The use of the standard sample size of 500 families, though clearly common, is far from universal. The average oblast' sample size was only 440 in 1978, and this average includes those oblasti with samples larger than 500, such as the sample of 1,200 in Moscow oblast' in 1968 (Table B3.7)[13]. It follows that many oblasti have samples smaller than 500, and indeed smaller than the minimum necessary size of 400 given by Barbashov et al (1968). It appears that samples of 500 are used in the Ukraine and probably in central Russia[14], but in Belorussia the average sample per oblast' is about 300 (1,580 for five oblasti in 1964). Even smaller samples are used in sparsely populated regions. Shvyrkov (1965) cites data relating to an unidentified "northern oblast'" from which we can infer a sample size of 240 families (using the method described in footnote 12). And oblast' samples in Kaazakhstan fall in the range 80-140 (Uchebnoe...1958 p.215).

If the view given in Barbashov et al (1968) about the necessary minimum sample size is generally accepted, one would expect survey data not to be used for calculations at oblast' level where the sample is below 400. That this is so is implied by advice on how to use survey data given in Uchebnoe... (1958 p.215). Errors are often made, we are told, through neglect of the rule that data should only be used within the limits of the populations which the survey is designed to describe. For these alone will checks of representativeness have been carried out. In the RSFSR and the Ukraine provision is made for collective farmer data to be representative at oblast' level, but this is not so in Belorussia and Kazakhstan, where the sample is representative only for the Republic as a whole (Konlov et al 1965). (Worker data are designed to be representative at the level of the industrial region, midway between oblast' and large Republic.) Thus an important function of

the survey - the provision of data for various calculations at oblast' level - can be fulfilled in only a proportion even of the oblasti which are covered by the survey.

7 The geographical distribution of the sample within individual territories

A number of critics of the FBS state that the geographical distribution of the sample within individual territories is often uneven. Kuznetsova (1978) writes that the sample "in some oblasti" is not representative of the oblast' as a whole, while according to Rimashevskaya (1968b) and Stolmov (1972) it is only in some oblasti that the sample is representative of the oblast' as a whole. The little evidence available on this question suggests rather mixed conclusions.

Let us consider first the survey of workers and employees. Apparently referring to the need to ease the burden of travelling imposed on interviewers, Postnikov (1953) mentions "the necessity for some concentration of (sampled) families in population points". Zhutovskaya (1966) links the uneven geographical distribution of the sample to the uneven pattern of representation of different types of enterprise in the sample (see Chapter B4). The FBS represents "mainly workers employed in the basic leading enterprises of the branches covered by the survey, situated as a rule in the large industrial centres". Most of these workers live in State flats, while workers in smaller centres tend to live in private houses, so that the geographical bias leads to an under-estimation of such expenditures as those on building materials and private-plot husbandry.

Some of the available descriptions of oblast' samples of workers and employees make a point of noting that not only the oblast' centre but also other towns in the oblast' are covered. Thus, Vladykin (1955) reports that the ten enterprises covered by the survey in Kuybyshev oblast' are situated in four different towns. Babayev (1972) tells us that, of the 425 families

of industrial workers covered by the FBS in Lvov oblast' in the Ukraine, 350 live in Lvov itself and 75 in the three next largest towns of the oblast'. The majority of descriptions of oblast' samples, however, include no information on the distribution of the sample among different towns. It appears likely that the sample of workers and employees is usually heavily concentrated in the oblast' centre.

For Latvia we have data which, while admittedly not conclusive, suggest to us a concentration of the urban sample in the Republican capital of Riga. In 1965 only 100-150 families of workers and employees were in the survey in the whole Republic (Rimashevskaya 1965). But ten years later the figure was 245 for Riga alone (Venetskii and Bauman 1975). The sample for Latvia had obviously been greatly expanded, perhaps in response to the complaint of Rimashevskaya. We presume that this expansion was part of the expansion of the sample of families of workers and employees for the whole USSR from 25,000 to 35,500 in 1969 (see Chapter B2), but Latvia's "fair share" of this expansion would have increased its sample only to about 150-200. However, a proportional representation of the urban population of the Republic in which Riga had a subsample of 245 would require a Republican sample of 500 - that is, it would imply an implausible fourfold expansion of the sample in Latvia[15].

Turning to the survey of collective farmers, the established procedure for the selection of collective farms is designed to provide for a fairly even distribution of the sample over the territory of the oblast'. When systematic selection of collective farms from lists results in two or more farms being taken from the same raion, all but one of these are to be replaced by farms from other raiony. The replacement should be adjacent to the original on the list, and therefore similar in economic orientation and per-capita income (Krylov 1957 p.80, Sergeev 1968 p.409)[16].

This rule should yield a fairly even territorial distribution in

oblasti where a considerable number of collective farms are selected, certainly in all those oblasti with collective farmer samples of 500 and over. It could not be relied upon in sparsely surveyed oblasti, and these tend to be the ones where geographical obstacles also militate against an even coverage.

Thus Tadzhikistan comprises two oblasti - Leninabadskaya and Kulyabskaya and a large area directly under Republican subordination. Collective farms are surveyed in each of these three territories, and the statistical handbook assures us that these are "evenly distributed over the basic agricultural regions of the Republic" (NKh TadzhSSR, 1961). However, these are not the only regions of the Republic, and Postnikov (1961) warns us that

> we must use (survey data) ... with great care. For example, the regions of Tadzhikistan vary greatly in climatic and economic conditions: conditions of life are very different in the lowlands and in the mountains. It is extremely complicated to organise a representative sample in the mountainous areas. Of course, survey data for Tadzhikistan scarcely describe the consumption of the population of the Republic as a whole.

The implied exclusion of the mountainous areas is consistent with the information in the handbook: in Kulyabskaya oblast' only the Kulyabskaya group of raiony are covered, while in the area under Republican subordination the Gissarskaya and Vakhshskaya valleys are covered.

Pechora raion of the Komi ASSR in the far north of European Russia provides a counter-example of the coverage of a far-flung area (Filippov 1961). Two collective farms are covered, situated at 85 and 287 kilometres respectively from the raion centre. The raion Inspectorate maintains communication with the interviewers by steamboat in summer and by aeroplane in winter. As this is the only report of its kind, we suspect that the situation it describes is exceptional.

Thus we can be confident that the sample is reasonably evenly distributed only as regards densely peopled rural areas.

8 Conclusion

The FBS is conducted in almost two-thirds of the oblast'-level territories of the USSR, although in only a little over one-quarter of the territories are both "workers and employees" and "collective farmers" covered. The temperate regions of the USSR, such as Central Russia, Belorussia and the Baltic Republics, are relatively intensely represented in the sample, while sparsely populated regions with extreme climates, such as the Far North and the mountainous areas of Central Asia, are very poorly represented. There is a tendency to use a standard oblast' sample size in the survey of collective farmers.

The distribution of the sample within oblast'-level territories is often uneven. In particular, workers and employees are sampled predominantly in the large industrial centres.

Notes to Chapter B3

1 A new oblast' in Uzbekistan was created in April 1982 (Radio Liberty Research RL 204/82 of 19.5.82).

2 This estimate may be inexact. Possibly there are a few other local Inspectorates involved in the conduct of the FBS. On the other hand, there were possibly some oblast' statistical administrations involved in the conduct of the FBS representatives from which did not attend this meeting.

3 The coverage of only 50 collective farmer families in East Kazakhstan oblast' in 1965 (Konovalov 1965) is not incompatible with this statement. Many collective farms in the Republic were in process of conversion into State farms at this time, and the combined sample of collective farmers and workers on State farms in East Kazakhstan oblast' was 75.

4 A small amount of FBS data appears unpredictably in Soviet statistical handbooks from time to time. Out of 480 statistical handbooks examined, pertaining to the USSR, particular Republics, oblast'-level territories and towns, 45 contained some FBS data, out of which twelve gave some methodological information about the survey.

5 These articles purport to describe the life of particular families, as revealed by their participation in the FBS. They are a useful source of information on the sample because they give the geographical location of the families. A more thorough search of the Soviet press would no doubt throw up some articles missed by us.

6 We consider such extrapolation from territories covered in 1939 to those covered since 1950 as fairly reliable for the following reasons: (i) there are many indications of the pre-war pattern being adhered to when the survey was reconstructed after the war, for example, in typical sizes of territorial samples of collective farmer families - as in Kazakhstan and the Ukraine; (ii) from Postnikov (1966) we know of 40 territories covered in 1939. We have independent evidence for the post-war coverage of 32 of these, and evidence against the post-war coverage of only one; we extrapolate for the remaining seven; and (iii) of the 30 territories identified as having a post-war sample of collective farmer families and which were Soviet in 1939, 27 were covered in 1939.

7 The Inspector of State Statistics of Pechora raion here describes how his Inspectorate started in 1960 to collect budgets of collective farmers in two of the six collective farms of the raion. Clearly it is impracticable to organise the survey directly from oblast' offices in sparsely populated far-flung regions; the exceptional nature of this example of the involvement of a raion office in the conduct of the FBS is therefore further proof of the poor coverage of such regions.

Moreover, Pechora raion is by no means one of the most isolated

areas of the USSR. Although it lies at a very northerly latitude, it is one of the three raiony out of the thirteen in the Komi ASSR which has a railway station on the single line running through the ASSR to Vorkuta (RSFSR - administrativno-territorialnoe delenie, M 1965, p.343).

Arkhangelsk oblast' in the Far North is covered by the survey, but only workers are surveyed there, who may well be all in the main town (see Section 7).

In view of the great degree of continuity in survey organisation over time, it is also relevant to note that the sample of collective farmer families in 1939 excluded the whole of the North, Far East and Siberia, with the exceptions of Novosibirsk oblast' and Altai krai (Postnikov 1966).

8 In some cases sample sizes are not given directly by the sources cited, but have been imputed from other information, such as the number of enterprises covered by the survey or the number of survey interviewers.

9 If we bring into the comparison the more tentatively estimated sampling fractions based on 1939 sample sizes (see Table B3.10), the range is even wider, extending up to 3.2 per 1,000 in rural Nikolayev oblast' (Ukraine).

10 One should not be misled by the fact that even relatively favoured local authorities, concerned with their own needs for data and under the influence of the Soviet belief in the need for very large samples, tend to complain that the sample for their area is too small. We have seen (Chapter B2) how Sukhova (Vsesoyuznoe... 1969 p.395) averred that the (unusually large) sample of 1,200 budgets in rural Moscow oblast' in 1968 was too small to give a true picture of the private plot economy. Morozov (1961) argues that coverage of "only" 0.14 per cent of collective farmer households in Georgia is quite inadequate to the needs of Georgian TsSU; the corresponding fraction for the USSR as a whole is 0.10 per cent.

11 From Shvyrkov's Table 7 one simply derives the relative sampling error E of expenditures on different items, and from Table 42 (p.105) their coefficient of variation V. The two Tables have in common a single item - milk - and this enables us to calculate the sample size n from V and E for milk:

$$n = (V/E)^2, \text{ as } E = 100\sigma/(\bar{x}.\sqrt{n}), \quad V = 100\sigma/\bar{x}.$$

12 This refers to variation in sampling fraction resulting from the use of standard sample sizes. Very low sampling fractions in oblasti where the sample is much smaller than the 500 standard (see below) may contribute appreciable sampling error to global data, even if bias is absent from sample selection. Corrective weights cannot, of course, be applied where there is no data at all - in the many oblasti where no collective farmer budgets are collected.

[13] Two textbooks give illustrative oblast' sample sizes of 757 and 1,125 (Table B3.13).

[14] We know that Krasnodar krai, RSFSR, had a sample of 550 in 1963.

[15] Estimated population figures for 1.1.1975 are:

795,800 for Riga (Riga, Riga 1976, p.14);

1,623,000 for the urban population of Latvia as a whole (NKh LatSSR v 1976 g, Riga 1977, p.7).

Thus Riga accounted for 49 per cent of the urban population of Latvia in 1975.

[16] However, Grankov (1955) says only that "as a rule" one collective farm is surveyed in any one raion, while Ananyeva (1964) says that one or two farms are surveyed. Nevertheless, we believe that the rule of one collective farm per raion is on the whole observed because it is consistent with general TsSU practice in sampling collective farms. Thus, in some oblasti a sample survey of electrified collective farms is conducted, and these are selected in such a way that any one raion Inspector is responsible for collecting data from a single collective farm; only by way of exception may two collective farms be covered in one raion (Grankov 1955). Clearly this helps to standardise the workload of different raion Inspectorates. This may be a consideration in the budget survey as well, as raion Inspectors take part in checking the work of the interviewers (see Chapter C1).

CHAPTER B4

BRANCH AND OCCUPATIONAL COVERAGE OF THE SAMPLE

1 Introduction

When assessing a survey organised on the branch principle, it is necessary to analyse not only the territorial coverage of the sample but also its coverage of the branches of the national economy. This is the main purpose of this chapter. Questions of the coverage of various staff and skill categories of workers and employees, and of various types of collective farmer households, are also dealt with.

The branches of the national economy covered by the survey of workers and employees, according to different sources, are shown in Table B4.1(a). Branches known not to be covered are listed in Table B4.1(b). The branch coverage of the survey of workers and employees is reviewed in Section 2: the coverage of manual workers by branch is dealt with in sub-section (a), the coverage of employees by branch and occupational group in sub-section (b) and the coverage of pensioners and other categories of the non-employed population in sub-section (c).

In view of the extreme paucity of data on sample sizes by branch and occupation, our discussion is concerned mainly with the identity of the groups covered. Available data do, however, permit us in Section 3 to compare sampling fractions for certain groups of workers and employees in the RSFSR.

On the whole little is known about the factors which underlie the extent of coverage of different branches and groups. The question of the coverage of workers on State farms is a partial exception in this respect, and is explored as a case-study in Section 4.

The overall pattern of the coverage of different groups of workers

TABLE B4.1(a)

LISTS GIVEN BY SOVIET SOURCES OF BRANCHES OF THE NATIONAL ECONOMY AND OF GROUPS OF EMPLOYEES COVERED BY THE FBS

Branches and groups	Source of list	Krapivina (1966)	Postnikov (1953)	Uchebnoe... (1958 p 184)	Kats (1960 p 173)	Matyukha (1969 pp
	Year to which list relates	1935-9	1953	1958	1960	1960
Workers in heavy industry						
Engineering and metalworking		*	*	*	*	*
Ferrous and non-ferrous metallurgy		*	*	*	*	*
Iron ore		*	*	*	*	
Coal mining		*	*	*	*	*
Oil		*	*	*	*	
Electricity		*	*	*	*	*
Chemicals			*	*	*	*
Rubber and asbestos		*	*	*	*	
Production of building materials (cement etc)			*	*	*	
Workers in light industry(1)						
Textiles		*	*	*	*	*
Garment industry(2)		*	*	*	*	*
Leather, fur and footwear		*	*	*	*	*
Glass, china and ceramics		*	*			

TABLE B4.1(a) (cont'd)

Branches and groups	Source of list	Krapivina (1966)	Postnikov (1953)	Uchebnoe... (1958)	Kats (1960)	Matyukha (1969)
	Year to which list relates	1935-9	1953	1958	1960	1960
Woodworking		*	*	*	*	
Paper		*	*	*	*	
Printing		*	*			
Food industry		*	*			
Workers outside industry						
Railway transport						*
Construction						*
State farms					*	*
Employees (including ITR)						
Of branches of industry covered		*	*	*	*	*
Of bodies of State and economic administration		*				
Of trade		*				
Teachers		*	*	*	*	*
Medical personnel		*	*	*	*	*

Notes: (1) There is good evidence that certain branches (for example, knitwear), omitted from t
Table because not given in any of the Soviet lists, are in fact covered (see text).

(2) Literally, "production of sewn products" (proizvodstvo shveinykh izdelii)

TABLE B4.1(b)

BRANCHES OF THE NATIONAL ECONOMY AND GROUPS OF WORKERS AND EMPLOYEES NOT COVERED BY THE POST-WAR FBS *

A. Workers in heavy industry

Fuel except coal and oil: gas, peat, shale, coke-chemical, production of artificial liquid fuel, other fuels

Production of thermal power

Extraction of non-ore deposits and production of abrasive, mica and graphite items

B. Workers in light industry

Timber

Tanning agent manufacturing

Fatty goods (candles, soap etc)

Medical industry (pharmaceuticals, instruments, glassware etc)

Production of cultural goods (film, musical instruments, drawing and stationery items, artistic items, toys)

Other branches of industry (eg combination fodder, water-mains)

C. Workers in particular categories of industrial enterprise

Enterprises of local and raion industry

Enterprises forming part of collective farms (eg food processing)

Very small or "petty" (melkie) enterprises

Craftsmen's cooperatives

D. Workers and employees in non-industrial branches of the national economy

Agriculture (except workers on State farms)

Rural intelligentsia (except teachers, doctors and nurses) - eg agronomists, veterinarians

Forestry

Transport (except railway workers) - ie water, motor, air, urban electric and other transport, and loading-unloading work

Communications

Agricultural procurements

Material-technical supply

Trade

Public catering

Everyday services (clothing repair, barbershops etc)

Residential services (probably)

Science and scientific services

The arts

Credit and insurance

State and economic administration

Other branches of the national economy (presumably including the Armed Forces, the KGB etc)

* Parts A and B of the list are compiled by eliminating branches known to be covered from the lists of branches of industry given by Savinskii (1954) and Yezhov (1967 pp 204-211). Part D is derived similarly from the list of branches of the national economy given in Trud v SSSR, M 1968. For discussion of non-coverage of non-industrial branches see Rimashevskaya (1965), Shvyrkov (1965), Karapetyan et al (1967) and Korovkin (1969). References for Part C are given in the text.

and employees in different territories is determined jointly by the pattern of territorial coverage and by that of branch coverage. This interaction of the two aspects of coverage is considered in Section 5.

In Section 6 we describe the exclusion from the survey of certain staff and skill categories of workers and employees. Section 7 deals with the exclusion of certain types of collective farmer households. In Section 8 we summarise the main findings of the chapter.

2 The branch and occupational coverage of the survey of workers and employees

In the absence of systematic accounts of the branch and occupational coverage of the survey of workers and employees, it is necessary to assess this coverage on the basis of disparate and sometimes mutually contradictory sources. The most important such sources are the various lists of branches allegedly covered (see Table B4.1(a)), a few published detailed breakdowns of oblast'-level samples and the comments of writers critical of the FBS[1].

Manual workers in the "basic" or "leading" branches of large-scale heavy industry, such as engineering, metallurgy and coal-mining, are preponderantly represented in the sample of workers and employees (Aleshina 1959, Matyukha 1960, Mochalov 1965 p.169, Rimashevskaya 1965, Kildishev et al 1980 p.384). There is on the whole a correspondingly poor representation of workers in light industry, small-scale industry and non-industrial branches of the national economy (transport, construction, agriculture etc.), and of employees of all categories. About one-half of all families of workers and employees stand no chance at all of inclusion in the FBS as a result of the non-coverage of the branches of the national economy in which their members work[2].

(a) The coverage of manual workers by branch

The heavy-industrial branches of engineering and metalworking[3], ferrous

and non-ferrous metallurgy, coal-mining[4] and oil are all well represented in the sample of workers. The sample of coal-miners in the RSFSR was reduced in the early 1960s, being considered larger than necessary, at the same time as the sample of workers in non-ferrous metallurgy was increased (Soveshchanie... 1964). The oil industry retains its traditional role as a locale for budget surveys (Matyukha 1967).

Other heavy-industrial branches such as electricity, chemicals, rubber-asbestos and the production of building materials (cement etc.), also seem to be reasonably well represented. The sample of workers in the building materials industry was increased in the early 1960s (Soveshchanie... 1964).

Although heavy industry taken as a whole is intensively covered, there are certain branches of it which appear not to be covered by the FBS, such as the extraction of various fuels (peat, shale, gas, artificial liquid fuel) and of non-ore deposits and the production of abrasive, mica and graphite items (see Table B4.1(b)).

Similarly, although light industry taken as a whole is poorly covered, there may be some exceptions. The textile industry figures in quite a few oblast' sample breakdowns, as does the garment industry. The textile industry is, like oil, a traditional locale of budget surveys[5]. The RSFSR sample of textile workers was even reduced as unnecessarily large in the early 1960s (Soveshchanie... 1964).

There is considerable evidence that the "leather, fur and footwear" branch is covered. The knitwear branch does not appear in lists of branches covered, but there is some local evidence for its inclusion.

The evidence with respect to a number of light-industrial branches is confused: "glass, china and ceramics", woodworking, paper, printing, the food industry. For example, Postnikov (1[illegible]) lists printing as a branch covered by the survey (Table B4.1(a)), but [illegible] Mints (1968) and Morovkin

(1969) complain of its exclusion. These branches, if not completely excluded, must be covered on a very small scale.

References to the coverage of the food industry are particularly inconsistent. Postnikov (1953) includes it in his list of covered branches. Rimashevskaya (1965) states that it is "almost not covered". Karapetyan, Rimashevskaya and Sidlyarenko (1967) mention the exclusion of the meat, milk and fish industries, implying that other sub-branches of the food industry may be covered. Korovkin (1969) states that the food industry is completely excluded. According to Soveshchanie... (1969), coverage of the food industry was newly introduced in 1968-9. As we shall see in Section 5, the branch is covered neither in Moscow, where survey coverage is in general good, nor in Moldavia, where far more workers are employed in it than in any other branch. On balance it seems that the food industry is at best very poorly covered - a serious deficiency in view of its importance.

A number of light-industrial branches are definitely excluded: toiletries and cosmetics, cultural products (film, stationery, musical instruments, toys etc.), consumer durables and so on (See Table B4.1(b)).

As Soloviev and Druker (1981) point out, workers in industry constitute only about one-third of all workers; workers in other branches of the national economy (transport, construction, agriculture etc.) are poorly represented. Since about 1960 workers in railway transport, in construction and on State farms have been included on a small scale, but other areas such as all non-rail transport are still excluded.

Railway and construction workers were included for the first time when the sample was expanded in 1960 (Krasnoshchekov 1962, Korovkin 1969). The samples of these workers were still very small in the 1960s (Tarfenova 1964), but were significantly expanded in 1968-9 (Soveshchanie... 1969).

Matyukha (1969) reports that the survey was extended to workers on State farms in 1960, though Dolgushevskii and Khristich (1976) state this

was done as early as 1957. Their coverage is confirmed by Kats (1960 p.173). The sample in the RSFSR was increased in the early 1960s (Soveshchanie... 1964), and coverage was significantly expanded in 1968-9 (Soveshchanie... 1969). The number of budgets covered, however, remains "small" or "insignificant" (Levin 1969 p.151, Levin 1973 p.254, Levin 1977, Soloviev and Druker 1981). (For further discussion see Section 4).

Various minor categories of industrial enterprise are also excluded from the survey: enterprises subordinated to raion or other local bodies ("local industry" and "raion industry"), and enterprises, such as food-processing plants, which form part of collective farms (Shvyrkov 1965, Korovkin 1969). Craftsmen working in cooperatives are excluded (Aleshina 1959).

"Petty" (melkie) enterprises are excluded from the survey irrespective of branch. According to Shvyrkov (1965) and Darbinyan (1971 p.87), enterprises with fewer than fifty workers are excluded. However, Krylov (1957 p.75), mentioning that TsSU instructions for 1952 recommend that selected small enterprises be replaced by larger enterprises adjacent on the list, gives an example in which an enterprise with 98 workers is replaced, suggesting that the cut-off point was then at least 100.

(b) The coverage of employees by branch and occupational group

Employees and engineering-technical personnel in industry are covered by the survey (Matyukha 1960, Parfenova 1964). As they are selected for the sample only in those branches of industry in which workers are selected, branch coverage is biased along broadly the same lines as in the survey of workers.

Apart from a single reference to the inclusion of employees at a construction administration in Minsk oblast' (Dedyulya 1972), there is no

evidence for the coverage of employees and engineering-technical personnel in such non-industrial branches of the productive economy as construction, transport and agriculture (State farms). Several critics note that, with the exception of a few rural teachers and medical personnel, the "rural intelligentsia" (for example, veterinarians and agronomists) is not covered by the survey (Karapetyan, Rimashevskaya and Sidlyarenko 1967, Levin 1969, Soloviev and Druker 1981).

The employees of "non-productive" branches of the national economy[6] are in general "almost not covered" (Rimashevskaya 1965), although their representation was significantly increased in 1968-9 (Soveshchanie... 1969). Of particular importance is the exclusion of wholesale and retail trade, in which more than five million employees work (Darbinyan 1971 p.87). Public catering, everyday services (laundries, barbershops, repair of clothing and footwear etc.), science and art, and State and economic administration are likewise excluded (Korovkin 1969)[7]. The evidence on the coverage of residential services is conflicting[8].

The only "non-productive" branches clearly known to be covered, though on a relatively small scale, are education and health care. Primary and middle school teachers, doctors and intermediate medical personnel (nurses etc.) are included (Matyukha 1960, Parfenova 1964). The representation of these groups was significantly increased in 1968-9 (Soveshchanie... 1969). Other groups of employees working in education (pre-school and higher educational establishments as well as libraries) and health care (auxiliary clerical and cleaning staff etc.) are not covered. Of the middle school teachers selected for the sample, one-third are taken from urban areas and two-thirds from rural areas, while of the medical personnel selected, two-thirds are taken from urban areas and one-third from rural areas. One-quarter of the medical personnel selected are doctors, and three-quarters are intermediate medical personnel. For reasons of convenience teachers

and medical personnel are selected only in population points where other survey participants are also resident (Krylov 1957 p.77)[9].

(c) The coverage of pensioners and other categories of the non-employed population

The evidence concerning the coverage of old-age pensioners up to the mid-1970s is conflicting. Aleshina (1959), Karapetyan et al (1967) and Levin (1969) all state that pensioners are excluded, but Ekonomicheskaya... (1970 p.278) states that they are included in the survey. Parfenova (1964) also reports a small sample of pensioners.

Whatever the situation at an earlier period, there is no doubt that pensioners have been covered in the last few years (Soloviev and Druker 1981, Dumnov, Pisarenko and Te 1981). N.M.Rimashevskaya dated the coverage of pensioners from about 1977 (Shenfield 1982b).

However, the sample of pensioners has been built up by retaining participants in the survey after their retirement instead of always dropping them from the sample on retirement. Thus biases of the survey of workers and employees are carried over into the survey of pensioners. For some time, moreover, the sample of pensioners will continue to be heavily biased in favour of relatively young, recently retired persons (Shenfield 1982b).

Apart from old-age pensioners, the non-employed population (the disabled, students etc.) is not covered by the FBS.

3 Comparison of sampling fractions for certain groups of workers and employees in the RSFSR

Some data on sample sizes for teachers, medical personnel, ITR and employees in industry, workers in construction and in railway transport and pensioners in the RSFSR, given by Parfenova (1964) and Sabinin (1966), enable us to calculate sampling fractions per 1000 persons of the workforce for some of these groups. The sample sizes are set out in Table B4.2, and

TABLE B4.2

RSFSR SAMPLE SIZES FOR VARIOUS GROUPS OF WORKERS AND EMPLOYEES

Groups	Sample sizes according to: Parfenova (1964)	Sabinin (1966)
1. Primary school teachers		
2. Middle school teachers		
3. Doctors		
4. Intermediate medical personnel (nurses etc)		
Sum of groups 1. - 4.	444	(515)
5. Engineering-technical personnel (ITR) in industry	(1723)	2000
6. Employees (excluding ITR) in industry	(692)	800
7. Workers in construction		
8. Workers in railway transport		
9. Old-age pensioners		
Sum of groups 7. - 9.	(1453)	(1685)
Sum of groups 5. - 9.	3868	(4485)
Sum of groups 1. - 9.	4312	5000
Sum of groups 1. - 6. (all employees in survey)	(2859)	(3315)

Note: Bracketed figures have been imputed as follows. We assume that the ratio of the figures given by the two sources for groups 1. - 9. taken together (5000/4312 = 1·16) applies to each group separately. Thus the sample for groups 1. - 4. is assumed to rise from 444 to 515. The figure for groups 7. - 9. in 1966 is obtained by subtraction, and then scaled down to the 1964 level. The figures for groups 5. and 6. are similarly scaled down.

TABLE B4.3

RSFSR SAMPLING FRACTIONS FOR VARIOUS GROUPS OF WORKERS AND EMPLOYEES (1)

Groups of workers and employees	RSFSR sample sizes		RSFSR population sizes		RSFSR sampling fractions	
	1964	1966	1964	1966	1964	1966
			'000	'000	per 1000	per 1000
Primary and middle school teachers, doctors and intermediate medical personnel	444	515	7075	7769 (2)	0·06	0·07
Workers in construction and in railway transport	1353	1585 (3)	4616	4787 (4)	0·29	0·33
All workers and employees in industry	9427 (5)		17242 (6)		0·55	

Notes: (1) The sample sizes in the first two lines are taken from Table B4.2. Population sizes are taken from the workforce data in Trud v SSSR, M 1968, pp 26-7, 42-3. It should be noted that the sampling fractions, calculated here per 1000 persons in the workforce, are not comparable with the sampling fractions, calculated per 1000 families of population, given for territorial units in Chapter B3.

(2) For 1966 we have the sizes of those parts of the health and education workforces covered by the FBS: "curative-prophylactic and sanitary-prophylactic establishments" (excluding sports etc) 4 199 000 + "primary and middle schools" (excluding pre- and post-school establishments, cultural establishments, children's homes etc) 3 570 000 = 7 769 000. For 1964 we have figures only for health and education as a whole, but we assume the same ratios between parts covered and not covered by the FBS as in 1966.

TABLE B4.3

Notes (cont'd): (3) From the figures for the sum of groups 7. - 9. in Table B4.2, we have here subtracted the estimate of sample size for pensioners of 100 (see Chapter B2).

(4) Sum of workforces in construction ("main activity" only) and in railway transport

(5) This estimate is obtained by applying to the figure of 10 000 workers in the RSFSR sample in 1963-4 (Shvyrkov and Shvyrkova 1973) an estimated ratio of "workers and employees in industry only" to "all workers" for the USSR as a whole in 1966-7. From the 15 700 workers in the USSR sample given by Matyukha (1967), we subtract 2000 workers on State farms and 1685 workers in construction and in railway transport (estimates for 1966) to obtain approximately 12 000 workers in industry. Adding 2800 ITR and employees in industry, we have 14 800 workers and employees (including ITR) in industry. The required ratio is therefore 14 800/15 700 = 0·94.

(6) Workers and employees in industry (including ITR; industrial-productive personnel only)

the calculations of sampling fractions are shown in Table B4.3, together with the corresponding sample fraction for "workers and employees in industry", introduced for purposes of comparison.

It is clear that, at least in the mid-1960s, teachers and medical personnel have been much less intensively represented in the FBS than workers and employees in industry, with workers in construction and in rail transport occupying an intermediate position.

4 Factors underlying the coverage of workers on State farms

We noted in Chapter B1 the tendency for the social, territorial and branch structure of the FBS sample to lag behind change in the structure of the population, as a consequence of the slow rate of renewal of the sample. There has been a long-term tendency in Soviet agriculture for State farms to replace collective farms. An examination of the reflection of this trend in the FBS sample will serve as a case-study of the general problem of structural lag[10].

In 1959 the survey covered 1000 families of workers on State farms (Kats 1960). By 1980-1 coverage had increased to 5800 families (Posobie... 1980, Dummov, Pisarenko and Te 1981; see Chapter B2). Although this is by no means an insiginificant level of representation[11], the intensity of coverage of families of workers on State farms has been and remains much lower than that of families of collective farmers. This is demonstrated by the comparison of sampling fractions for State farms and for collective farms, for various territories at various times and for the USSR as a whole in 1980-1, shown in Table [illegible]. The rise in the relative weight of workers on State farms in the sample, considerable as it has been, nevertheless lags behind the trend from collective farms towards State farms in the population.

The [illegible] of taking proper account of the replacement of collective by State farms is impeded by the way the FBS is organised. The Budget Survey

TABLE B4.4

COMPARISON OF FBS COVERAGE OF COLLECTIVE AND OF STATE FARMS

Area (source, year)	Collective farms			State farms			Ratio of coverage of collective farms to coverage of State farms
	Sample[1]	Pop'n	Coverage %	Sample[1]	Pop'n	Coverage %	
Krasnodar krai, RSFSR (Sidlyarenko 1964)	22	338	6·5	3	190	1·6	4·1
Minsk oblast', Belorussia[3] (Dedyulya 1972)	13	?	?	2	?	?	2·5
Odessa oblast', Ukraine (Klebanov 1968)	20	426	4·7	1	103	1·0	4·8
Lvov oblast', Ukraine (Babayev 1972)	20	412	4·9	1	52	1·9	2·5
East Kazakhstan oblast', Kazakhstan (Konovalov 1965)	2	30	6·7	1	60	1·7	4·0
Moldavia (NKh MSSR v 1960 g)	21	557	3·8	0	61	0·0	inf.[4]
Tadzhikistan (NKh TadzhSSR v 1961 g)	18	351	5·1	0	41	0·0	inf.[4]
USSR (Posobie... 1980, Dumnov, Pisarenko and Te 1981)	908	26000[5]	3·5	232	20767[5]	1·1	3·1

Pop'n - Population

Notes

(1) In some cases the number of budgets collected has been converted into the number of collecti or State farms by dividing by 25, on the well-supported assumption that one interviewer cove one collective or State farm. Thus, from 22700 collective farmer budgets and 5860 State farm worker budgets we obtain for the USSR in 1980-1 908 collective farms and 232 State farms.

(2) The ratio is calculated without taking into account the relative sizes of collective and Sta

TABLE B4.4

Notes (cont'd)

than collective farms, the ratios given understate the over-representation of collective farmers compared to workers on State farms.

(3) Population figures for Minsk oblast' are not available. The coverage ratio is based on population figures for Belorussia as a whole. The ratio of collective to State farms in Belorussia in 1972 was 2·6, compared to a ratio in the FBS sample for Minsk oblast' of 6·5.

(4) The ratio is infinite because State farms are excluded altogether from coverage.

(5) The population figures for the USSR relate to the end of 1979. These are the most recent data available.

Departments of TsSU at oblast', Republic and All-Union levels are organised in two separate hierarchies, one dealing with budgets of workers and employees and the other dealing with budgets of collective farmers (see Chapter A2). Each of these two sectors collects, processes and aggregates budget data independently. Aggregated data are submitted upwards from level to level within each sector[12]. Workers on State farms belong to the "social group" of "workers" (see Chapter B2), and data relating to them are collected and processed by the Departments dealing with budgets of workers and employees. When collective farms covered by the FBS are converted into State farms, corresponding adjustments in the organisation of the two sectors are required.

This problem was raised at the conference of statisticians in 1957 by T.U.Uvashev, the Head of TsSU Kazakhstan (Vsesoyuznoe... 1959 p.90). It will be seen from Table B4.5 that the replacement of collective by State farms has proceeded particularly rapidly in Kazakhstan. Uvashev reports that he had unsuccessfully proposed to TsSU USSR that, in view of the declining number of collective farmers in the Republic, the Budget Survey Departments for collective farmers' budgets in the oblast' statistical administrations be merged with the Budget Survey Departments for workers' budgets:

> Questions regarding the structure of statistical bodies are decided by TsSU (USSR) without regard to our proposals (ie. the proposals of TsSU Kazakhstan - SDS), and incorrect decisions do great damage to our work...
>
> (Uvashev here gives an example relating to the organisation of the population census - SDS).
>
> ...TsSU (USSR) did not accept our proposal to merge Departments and Sectors of workers' budgets and of collective farmers' budgets. As a result of the recent transformation of collective farms into State farms, in those areas where formerly collective farmers' budgets were collected not one collective farm remains and only budgets of workers on State farms are collected. Nevertheless, there exist as before Sectors of Collective Farmers' Budgets in the oblast' statistical administrations. It seems that some leading officials of TsSU (USSR) do not want to notice what is happening in practical life: a structure set up at some point in time appears ideal to them, and it is not so at all.

TABLE B4.5

CHANGE IN THE RATIO OF COLLECTIVE TO STATE FARMS IN KAZAKHSTAN

Year	Number of collective farms '000	Number of State farms '000	Ratio of number of collective farms to number of State farms
1953	3·1	0·3	10·3
1956	2·7	0·6	4·5
1957	1·8	0·8	2·3
1960	1·4	0·9	1·5
1965	0·5	1·5	0·3
1970	0·5	1·6	0·3
1978	0·4	2·0	0·2

Sources: Figures for 1956 and 1957 from Vsesoyuznoe... (1959 p 90). Remaining figures from the annual statistical handbooks Narodnoe Khozyaistvo Kazakhskoi SSR.

The nature of the procedures associated with the out-of-date organisational structure and doing "great damage" to the FBS cannot be inferred with any certainty from Uvashev's speech. The centralisation of decision-making in the hands of TsSU USSR presumably extends to the determination of staffing complements for the Departments of all TsSU offices, for otherwise TsSU Kazakhstan could solve its problem by reducing the superfluous Departments to some nominal level. Rigidity in staffing complements would help account for lags in adjusting sample structures to the changing balance between collective and State farms in the population of such territories as East Kazakhstan oblast' (see Table B4.4).

However, Uvashev is referring to territories in which collective farms have completely disappeared. With what work are oblast' Sectors of Collective Farmers' Budgets in this case occupied? Unless they do nothing (which would cause no damage apart from waste of resources), they must continue to collect budgets from families which used to be families of collective farmers but are now families of workers on State farms. The question then arises: to which Department at the next higher level - that is, at the level of TsSU Kazakhstan - do they submit the data aggregated by them? If they submit the data to the Department of Budgets of Workers and Employees for inclusion in the totals for budgets of workers on State farms, then the only damage caused is some confusion in the formal organisational structure. If "great damage" is done, this probably means that the data are submitted, as before, to the Department of Budgets of Collective Farmers for inclusion in the totals for budgets of collective farmers. It follows that the sample of "collective farmers" at Republic and All-Union levels in fact includes a certain proportion of workers on State farms, while the sample of "workers on State farms" takes account of only a proportion of the workers on State farms actually covered. This would affect the reliability of FBS data ostensibly relating to these groups[13].

5 The interaction of territorial with branch coverage

In Chapter B3 we examined the uneven territorial coverage of the sample of workers and employees. In this chapter we have examined its uneven coverage by branch and occupational group. It remains to consider how territorial coverage interacts with branch and group coverage to determine the detailed pattern of coverage by territory, branch and group represented by the sample quotas set by TsSU USSR (Chapter B1). The data available on this pattern of coverage are extremely incomplete. They make possible only a comparison of the branch coverage of workers in industry in five areas and some comments on the territorial distribution of the sample of teachers and medical personnel. Conclusions are accordingly tentative.

The five areas for which we have information on the branch coverage of workers in industry are the cities of Moscow and Tbilisi (capital of Georgia) and the Republics of Moldavia, Azerbaijan and Tadzhikistan. All the information relates to the period 1956-61. The data are set out in Table B4.6[14].

The comparison of the five areas clearly confirms the known concentration of the sample of workers in industry on the "basic industrial regions" in the centre of the USSR (Aleshina 1959). The branches covered in Moscow account for about three-quarters of the industrial workforce of the capital, while the branches covered in Tbilisi account for little over one-half of its industrial workforce and the branches covered in Moldavia for only one-third of its industrial workforce[15]. Coverage in the central areas also appears to encompass a broader range of branches than in the peripheral areas: ten branches are covered in Moscow, five in Tbilisi, Moldavia and Azerbaijan and only three in Tadzhikistan.

On the evidence of this comparison, the territorial coverage of workers in heavy industry seems to be more uneven than that of workers in light industry. Some favoured branches of light industry - textiles, the garment

TABLE B4.6

COVERAGE OF WORKERS IN INDUSTRY BY BRANCH IN FIVE AREAS[1]

Key:
- * branch covered by the FBS
- (*) branch covered by the FBS even though it accounts for a very small[2] proportion of the workforce in the area
- x branch not covered by the FBS even though it accounts for a substantial[2] proportion of the workforce in the area
- (x) branch not covered by the FBS, but it accounts for a very small proportion of the workforce in the area
- \- branch not represented in the area
- -? not clear whether branch represent in the area

Branch of industry	City of Moscow	City of Tbilisi, Georgia	Moldavia	Azerbaijan	Tadzhikistan
	1956	1960	1960	1960	1960-1
Heavy industry					
Engineering	*	*	x	*	x
Metalworking	*	x	x	x	*
Ferrous metallurgy	*	-	-	x	-
Non-ferrous metallurgy	-?	-	-	x	-
Coal-mining	-	-	-	-	x
Oil	-	-	-	*	-
Electricity	-	(*)	(x)	*	x
Chemicals	*	(x)	(x)	x	x
Rubber and asbestos	*	(x)	-?	-?	-?
Production of building materials (cement etc)	x	x	-	-	

TABLE B4.6 (cont'd)

Branch of industry	City of Moscow	City of Tbilisi, Georgia	Moldavia	Azerbaijan	Tadzhikistan
	1956	1960	1960	1960	1960-1
Light industry					
Textiles	*	*	*	*	*
Garment industry	*	*	*	*	*
Leather and fur	*	(x)	*	-?	-?
Footwear	*	*	*	-?	-?
Knitwear	*	-?	*	-?	-?
Glass, china and ceramics	-?	-?	(x)	-?	-?
Woodworking	x	x	x	x	x
Paper	-	(x)	x	x	-?
Printing	x	x	-?	-?	-?
Food industry	x	x	x	x	x
Proportion of workers in industry in the area accounted for by the branches covered	70-80 %	56 %	34 %	Not known(3)	Not known(3)

TABLE B4.6

Notes:

(1) The sources used are listed in footnote 14 to this chapter.

(2) It is useful to distinguish between the case in which a particular branch of industry accounts for a very small proportion of the industrial workforce in an area and the case in which it accounts for a considerable proportion of the workforce. The coverage of a branch which accounts for a very small proportion of the workforce in an area suggests that the general coverage of that branch throughout the USSR may be very high, while its non-coverage in that area may have little significance for its general coverage throughout the USSR.

We take two per cent as the cut-off point between a "very small" and a "considerable" proportion of the workforce. In Moscow all branches account for considerable proportions of the workforce. The distinction cannot be made for Azerbaijan and Tadzhikistan, as the branch breakdown of the workforce for these areas is not available.

(3) The proportion cannot be calculated because the branch breakdown of the workforce is not available.

industry, footwear and knitwear - are covered in all of the areas considered in which they exist, while other branches of light industry - woodworking, paper, printing and the food industry - are covered in none of the areas considered. Of the branches of heavy industry, engineering, metalworking, ferrous metallurgy, electricity and chemicals are covered in some of the areas in which they exist but not in others[16]. In none of the areas is the production of building materials covered. Coal-mining is not covered in Tadzhikistan, the only one of the five areas in which it exists, although we know that it is well covered elsewhere[17].

Thus, the over-representation of heavy industry in the sample of workers at All-Union level notwithstanding, heavy industry may be under-represented in particular territories. Moldavia provides the most striking example of this phenomenon: there is a substantial heavy industry in the Republic, but the FBS covers there only a number of branches of light industry[18]. The pattern of coverage is evidently a complex one which resists attempts at simple explanation.

Parfenova (1964) provides a little information about the territorial coverage of the sample of teachers and medical personnel in the RSFSR in 1963. The 444 surveyed families belonging to these groups were very unevenly distributed over about one-quarter of the oblast'-level territories of the RSFSR (18 out of 71 territories). While the average number of families surveyed per territory covered was about 25, considerably fewer families were surveyed in some territories (for example, ten families in Tula oblast' and seven in Rostov oblast'). Table B4.7 shows the lack of any clear connection between sample sizes and the populations of territories.

6 The coverage of workers and employees by staff and skill category

Quite apart from the very partial coverage by territory and branch of the survey of workers' budgets, a very large proportion of workers in

TABLE B4.7

COVERAGE OF TEACHERS AND MEDICAL PERSONNEL IN INDIVIDUAL TERRITORIES OF THE RSFSR

Territory	Number of teachers and medical personnel in the FBS sample 1963	Population of territory millions	Number of teachers and medical personnel in the FBS sample per million population
Tatar ASSR	17	3·2	5
Tula oblast'	10	1·9	5
Rostov oblast'	7	3·9	2

Source: Parfenova (1964)

industry and construction are excluded from coverage as a result of belonging to various excluded staff categories. Employees are affected by some of these exclusions, though to a much lesser extent. The most important exclusion is that of the unskilled grades of manual worker.

First, a Soviet enterprise or organisation has a list of permanently employed staff (spisochnyi sostav rabotnikov), and samples of workers and employees for the survey are drawn only from this list. A number of minor categories of staff are thereby excluded, though there is no reason to believe that they account for a substantial proportion of the workforce:

i. staff on temporary hire for occasional one-off pieces of work unconnected with the basic activity (see below) of the enterprise - for example, consultancy, artistes' appearances, repair of inventory, whitewashing and painting - and paid out of the wage fund for non-list (nespisochnyi, neshtatnyi) staff;

ii. pupils of technical colleges doing their production practice at the enterprise;

iii. staff on secondment to study, on grants paid by the enterprise;

iv. staff on business missions (komandirovki) whose wages are not being retained for them at their place of basic work;

v. staff on unpaid leave in the slack season (where work is seasonal);

vi. replacement staff (sovmestiteli); and

vii. staff occupying more than one post (Posobie... 1980 p.307, Nazarov, Parteshko and Rumyantsev 1981 p.33).

Second, the list staff of an enterprise are divided into two categories: those employed on the "basic activity" of the enterprise, and those employed on "non-basic activity". The lists from which samples are drawn for the survey consist of staff on basic activity only (Nikishev et al 1980 p.384).

In an industrial enterprise, staff on basic activity are also called "industrial-production (promyshlenno-proizvodstvennyi) staff". They are all those "directly taking part in the process of production, in serving or managing it" (Nazarov, Parteshko and Rumyantsev 1981 p.38). This includes transport which serves production, stores, laboratories, data-processing facilities and administration as well as basic and auxiliary workshops. The staff on non-basic activity who are excluded are those in non-industrial units financed by the enterprise:

i. units providing social services to the workforce - pioneer camps, clinics, crèches, houses of rest, sporting facilities, libraries, clubs, canteens, accommodation services, the factory paper etc;

ii. the Department of Capital Construction, which repairs buildings and equipment; and

iii. subsidiary agricultural and other non-industrial enterprises.

In a construction organisation, it is construction and reconstruction work which comprises the basic activity,while subsidiary and service units do non-basic work. On State farms, staff on basic activity are those directly concerned with the production of agricultural products, while staff on non-basic activity are those employed in subsidiary industrial enterprises as well as service units.

Staff on non-basic activity appear to correspond to staff falling outside the system of tariff grades (Baklanov, Adamov and Ustinov 1970); these comprised 10 per cent of manual workers in industry and construction in 1965 (Vestnik statistiki 1966, 3. .

Third, industrial-productive staff comprise, besides the three major categories of manual workers, employees and engineering-technical personnel, two minor categories: (i) apprentices; and (ii) junior service staff and security staff. These two categories are certainly excluded from the survey.

Together they accounted for 4.5 per cent of all industrial-productive staff in 1956, 3.2 per cent in 1968 and 2.7 per cent in 1979 (NKh SSSR v 1956 g, p.50; NKh SSSR v 1968 g, p.205; NKh SSSR v 1979 g, p.147).

Finally, and most crucially, we must consider which grades of industrial-productive workers are included in the lists from which samples of workers are drawn for the survey. Most branches use a six-grade scale, some (for example, ferrous metallurgy and chemicals) a seven-grade scale, and a few use eight or even ten-grade scales. High grade numbers signify high skill levels. Matyukha (1967 p.17) states that samples of workers in an enterprise are drawn from two separate lists, one of "qualified" (kvalifitsirovannye) workers and one of "little qualified" (malokvalifitsirovannye) workers. He gives as his example the engineering industry, where a six-grade scale is in operation, grades 1-3 corresponding to "little qualified" and grades 4-6 to "qualified". Thereby he gives the impression that all grades are covered, though he does not explicitly state that this is in general so. We shall argue that this impression is misleading.

There seems to be no generally agreed system of classifying workers' "qualification levels" and relating them to grades in Soviet usage. The usage found in various statistics textbooks is shown in Table B4.8. Seven different terms have been found: "highly qualified", "qualified", "medium qualified", semi-qualified", "little qualified", "weakly qualified" and "non-qualified". ("Lowly qualified" is probably synonymous with "weakly qualified".) Most authors distinguish three qualification levels, selecting three of the terms from those mentioned. Only Kats (1957) gives a fourfold classification, for "certain [illegible]" only, as an alternative to the threefold classification. No author consulted uses a twofold classification. If, as Matyukha implies, this is the standard practice in the engineering industry, engineering is likely to be exceptional in this regard.

TABLE B4.8

TERMINOLOGY USED IN STATISTICS TEXTBOOKS FOR SKILL LEVELS OF WORKERS

Terms used	Reference no. (1) Year 1957 (a)	(1) 1957 (b)	(2) 1957	(3) 1966	(4) 1971	(5) 1978	(6) 1978	(7) 1979
Highly qualified (vysokokvalifitsirovannye)	*				*	*	*	*
Qualified (kvalifitsirovannye)		*	*	*	*	*	*	*
Medium qualified (srednekvalifitsirovannye)	*							
Semi-qualified (polukvalifitsirovannye)		*						
Little qualified (malokvalifitsirovannye)	*		*	*				*
Weakly qualified (slabokvalifitsirovannye)			*	*			*	
Non-qualified (nyekvalifitsirovannye)	*	*			*	*		

References

(1) Kats Ya D, Promyshlennaya statistika na predpriyatiyakh chernoi metallurgii, M 1957: two alternative classifications given

(2) Yezhov A I, Statistika promyshlennosti, M 1957

(3) Yezhov A I, Statistika promyshlennosti, M 1966

(4) Ekonomicheskaya statistika (eds. M I Leshchinskii and N N Ryauzov), M 1971

(5) Ekonomicheskaya statistika (eds. M G Nazarov and N N Ryauzov), M 1978

TABLE B4.8 (cont'd)

References (cont'd)

(6) Kurs statistiki promyshlennosti (eds. I G Malyi and V I Siskov), M 1978
(7) Yeremina N M and Marshalova V P, Statistika truda, M 1979

Shvyrkov (1965 Ch.2) mentions three qualification levels which correspond to those given by Yezhov (1957, 1966): "qualified" for grades 5 and 6, "little qualified" for grades 3 and 4, and "lowly qualified" for grades 1 and 2. Lists are compiled for the first two of these three groups, but "lowly qualified" workers are excluded from the survey. Their exclusion is confirmed by Darbinyan (1971 p.87). This explains the discrepancy between the use of only two lists in sample selection and the prevalence of classifications into three qualification categories.

The proportion of all workers in the excluded staff and skill categories, though very substantial, has fallen somewhat over time. In the 1950s only about 50 per cent of workers belonged to the higher and middle skill grades; the proportion had risen to almost 70 per cent by 1969 (Problemy... 1973 p.122). The proportion of unskilled and ungraded workers, excluded from the FBS, had therefore fallen from about 50 per cent to a little over 30 per cent. In 1965 the lower skill grades and the ungraded categories comprised 40 per cent of workers in industry and 35 per cent of workers in construction (Krevnevich 1971 p.134).

One can only speculate about the possible reasons why unskilled workers are not covered by the FBS. The practice may well originate in a judgement made in the 1930s that it was more important to monitor the living standards of skilled workers, whose morale was most crucial to the success of industrialisation. It must have been much easier to get better educated workers to keep proper budget records, and this may still remain a relevant consideration. Finally, the resulting upward income bias in the sample may be politically welcome.

7 Coverage of collective farmer families by type of household

Territorial biases aside, the survey of collective farmers suffers from fewer and less serious exclusions than the survey of workers and em[illegible].

so that the sample is probably more representative of the population as a whole. However, we have evidence for the exclusion of three categories of household: one-person households, households without able-bodied members and households which share a cow with other households. These categories must overlap to a great extent; many elderly people must be excluded from the survey as a result. The combined effect is to bias the sample towards the more prosperous households.

Unlike the survey of workers and employees, in which some one-person households as well as families are surveyed, the survey of collective farmers covers families only (Kildishev et al 1980 p.383). Shvyrkov (1965) informs us of the exclusion of households lacking able-bodied members. Krylov (1957 p.82) gives an illustrative table listing collective farmer households for the purpose of sample selection. One column is headed "note of absence of able-bodied persons and of selection"; in this column some households are labelled "selected" (for the survey) and others "no workdays". Although this is not made explicit in the text, it seems reasonable to infer that these are alternative annotations and that households which have contributed no labour to the collective farm are ineligible for inclusion in the survey.

Dolgushevskii and Khristich (1976 p.342) tell us that, before selecting households for the survey, all households of a collective farm are divided into two groups: those possessing private cows and those not possessing them. This would seem to ensure proper representation of both better-off and worse-off households. However, we realise from Shvyrkov (1965) that these two groups are not exhaustive; there is an intermediate group of households who share livestock with other households, usually between two families. This group is not covered by the survey. Shvyrkov believes that the exclusion introduces a significant bias towards better-off households.

Moreover, one textbook account of the survey implies that families without cows, as well as those sharing cows, are excluded. Kozlov, Ovsienko and Smirnskii (1965 pp.230-1) illustrate the method of selecting a cluster of 25 families in one collective farm as follows: "Suppose, for example, that there are 100 families possessing cows in a collective farm. We must select 25 households...".

Thus households which share a cow are definitely excluded from the survey, and it is possible that households without cows are also excluded. To assess the seriousness of the resulting bias, we need information on the distribution of cow ownership among rural households.

Masherov (1978) reveals that 32 per cent of rural households in Belorussia have no cows. For the USSR as a whole the situation is even worse. The rural population as of 1 January 1981 is estimated at 97.7 million (NKh SSSR v 1980 g, M 1981, p.7); as the average size of a rural family is 4.0 persons (Roganova 1976), there are 24.4 million rural households. There are only 13.2 million cows in private possession (NKh... p.245), or 0.54 of a cow per rural household, assuming urban households own no cows. As households in nomadic regions may own several cows each (Symons 1972 p.11), at least one-half of rural households must be without their own cow.

We have not been able to find data on the proportion of households sharing cows, but it must be considerable if we assume a fairly continuous income distribution.

Thus the bias resulting from the exclusion of certain types of collective farmer households must be substantial[19].

Conclusions

The branch and occupational coverage of the sample of workers and employees is very uneven. Heavy industry is on the whole well represented,

light industry less well so, with the exception of certain branches such as textiles. The very poor or non-existent coverage of the food industry is a serious deficiency. Rail (but not non-rail) transport and construction are moderately well covered, as are workers on State farms, although the balance between collective farmers and workers on State farms in the sample lags behind the long-term replacement of collective by State farms in the agricultural economy, in part as a result of the rigid departmental organisation of the FBS. Local industry and very small enterprises are not covered.

"Non-productive" branches of the national economy, such as trade and State administration, are not covered by the survey, with the exception of education and health-care. Old-age pensioners have been covered in recent years, but other non-employed groups (the disabled, students etc.) are still excluded.

The interaction of territorial with branch and occupational coverage is a complex one. The territorial coverage of heavy industry appears to be more uneven than that of light industry: there are areas (such as Moldavia) in which the survey covers certain branches of light industry while excluding heavy industry. The territorial distribution of the relatively small samples of teachers and medical personnel is also uneven.

Various staff and skill categories of workers and employees are excluded from the sample. The most serious exclusion is that of unskilled workers, who constituted about 30 per cent of the industrial workforce in 1970.

Certain types of collective farmer households are also excluded from the survey: one-person households, households lacking able-bodied members or contributing no labour to the collective farm, households sharing the possession of a cow with other households and possibly households possessing no cow. The effect of these exclusions is a bias in favour of better-off households.

Notes to Chapter B4

1 Let us cite two examples of detailed oblast'-level breakdowns. Vladykin (1955) informs us that 255 workers' budgets are collected in Kuybyshev oblast' from ten enterprises of four branches, including three engineering and two chemical enterprises. Venetskii and Matyukha (1968) give a branch breakdown of the 720 workers' budgets in an unidentified oblast': very heavy representation of textiles (429) and engineering (212), with some budgets from ferrous metallurgy (16), electricity (17), non-ferrous metallurgy (14) and chemicals (32).

A less important source of information is the identity of "typical" branches used in illustrations of the method of sample selection and the like.

2 Although one half is of the right order of magnitude, the sources are not clear or consistent enough to be at all precise. Rimashevskaya (1968) claims that the included branches cover only about 40 per cent of all families (of workers and employees). But elsewhere (Rimashevskaya 1965) she reports that the incomes survey showed that "almost half" of families had heads working in the excluded branches. As only one family member has to work in an included branch to give the family a chance of inclusion in the survey, this implies that over half of families are covered. Korovkin (1969) states that about 40 per cent of workers and employees are employed in the excluded branches, implying that over 60 per cent of families are covered. A family might be considered "covered" either if its head works in an included branch or if any member works in an included branch, and definitions to resolve the ambiguity are not given: this may explain the discrepancies.

3 Matyukha (1967), explaining the method of sample selection, "supposes" that 3700 out of 15700 workers' budgets are taken from the engineering and metalworking branch. If this figure is not in fact hypothetical, the branch is somewhat over-represented in the sample, accounting for almost 24 per cent of the worker sample as against only 19 per cent of the worker population (as estimated from data in Trud v SSSR, M 1968, pp.86-9).

4 Golovach (1963) mentions that metallurgy and coal-mining predominate in the Ukrainian sample.

5 Three of the six main pre-revolutionary surveys of workers' budgets - those of Davidovich in 1908, Shaposhnikov in 1909 and Goritskii in 1911 - were conducted in the textile industry (Matyukha 1967 pp.174-5). The importance of tradition in the coverage of the budget survey is exemplified by the continuity of observation at one of the oldest textile mills in Russia (founded in 1[illegible] Pavlovo-Posad, Moscow oblast', for which Vasilyeva (1965) is able to compare workers' budgets in 1909 and 1961.

6 In referring here to "non-productive" branches, we follow the loose usage of Soviet writers on the FBS, which does not exactly correspond to the conventions of Soviet national income accounting.

7 The authors of a joint Soviet-Czech work on budget surveys, Krutikov et al (1981 p.178), state that "highly placed persons, scientific and artistic workers" are not covered by budget surveys. This may be a general feature of East European budget surveys.

8 According to Kurs... (1961 p.433), the "residential services" branch is covered. Karapetyan et al (1967) and Korovkin (1969) state that it is excluded. According to Soveshchanie... (1969), it was one of the branches newly introduced into the survey in 1968-9, while according to Razvitie... (1971) its inclusion is only "envisaged".

9 It is curious that coverage of employees and ITR was broader before the war (we have information on the period 1935-9 from Krapivina 1966) than it has been since the war. In 1958 the sample - 4700 (Uchebnoe... 1958) - remained much smaller than the 8000 it had reached in 1940 (Matyukha, Postnikov and Samoilov 1958), though it had surpassed the 1940 level by 1967 (Matyukha 1967). Before the war coverage included several groups now excluded - employees and ITR employed by State and economic bodies, in trade, pre-school and higher education and in libraries - as well as the groups now covered.

10 The inflexibility revealed by this particular case-study may well be of an exceptionally extreme kind.

11 The statements by some critics of the FBS to the effect that coverage of workers on State farms is "insignificant" (Levin 1969, 1973, 1977) or "on a small scale" (Soloviev and Druker 1981) are rather exaggerated, possibly reflecting the common Soviet prejudice against samples of moderate size (see Chapter A4). This tendency of some writers to dismiss any but large samples as "insignificant" may help to explain the discrepancies among different sources regarding branch coverage (Section 2(a) of this chapter).

12 Until 1977, when unified budget survey forms were introduced, the two sectors of the Budget Survey Departments used different forms to collect and process budget data (Ekonomicheskaya... 1983 p.378).

13 If this confusion also affects the reckoning of the sizes of samples of families of collective farmers and of workers on State farms, then the under-representation of families of workers on State farms in the FBS may in reality be less severe than it appears.

14 Section 5 and Table B4.6 are based on the following sources:

(a) On the sample in Moscow in 1956: Moskva - razvitie khozyaistva i kultury goroda, M [illegible]:

(b) On the sample in Tbilisi in 1960: Tbilisi k 40-letiyu sovetskoi vlasti v Gru[illegible]. Tbilisi 1961;

(c) On the sample in Moldavia in 1960: Narodnoe Khozyaistvo Moldavskoi SSR v 1960 g, Kishinev 1961;

(d) On the sample in Azerbaijan in 1960: Razvitie Narodnogo Khozyaistva Azerbaidzhanskoi SSR i rost materialnogo i kulturnogo urovnya zhizni naroda, Baku 1961; and

(e) On the sample in Tadzhikistan in 1960-1: Narodnoe Khozyaistvo Tadzhikskoi SSR v 1960 g, Dushanbe 1961; Narodnoe Khozyaistvo Tadzhikskoi SSR v 1961g, Dushanbe 1962.

15 Calculations are based on data from statistical handbooks.

16 Table B4.6 may give the impression that the samples of workers in branches of heavy industry are more restricted geographically than is actually the case. Thus, Rabochii... (1969 p.231) refers to coverage of the engineering and metalworking branch in Leningrad, Moscow, Gorkii oblast' and Novosibirsk oblast', to coverage of the metallurgical industry in the Urals etc. Azerbaijan is of course the only one of the five areas covered by the Table in which oil is extracted.

17 Areas in which coal-mining is known to be covered are the Ukraine (Golovach 1963) and Karaganda oblast' in Kazakhstan (Rabochii... 1969 p.231).

18 There were about 105,000 workers in industry in Moldavia in 1960. Of these, about 30,000 were in heavy industry (mainly in engineering, metalworking and the production of building materials), about 36,000 in those branches of light industry covered by the FBS in Moldavia (textiles, the garment industry, leather, fur and footwear, and knitwear), and about 39,000 in those branches of light industry not covered by the FBS in Moldavia (mainly in the wood, paper and food industries).

19 Another bias in the survey of collective farmers arises from the regulation which lays down that, when a participating family divides into two with both new families remaining on the same collective farm, the sample retains only that family the members of which earn the larger number of labour-days per year (DePauw 1965 p.11).

CHAPTER B5

THE EXPERIENCE OF PARTICIPATION IN THE FBS AND ITS CONSEQUENCES

1 Introduction

In this chapter we are concerned with the way the representativeness of the sample and quality of the data collected are affected by the nature of the experience of participation in the FBS for Soviet families. In Section 2 we describe what participation involves for a family, emphasising the burden which it imposes on their time and effort. In Section 3 we consider the incentives and pressures which are brought into play to encourage continued participation in the survey. We are then in a position to discuss the puzzling question of the frequency with which families refuse to participate, whether from the outset or later on, in Section 4. In Section 5 we contrast two different strategies which families may use to cope with the burden of participation. In Section 6 we conclude by reviewing the harmful effects that the burden, and the incentives and pressures necessitated by it, have on sample representativeness and data quality.

These are politically sensitive issues - not only in their implications for data quality but also in their connection with the difficulties of relationships between citizens and the authorities. Information on the crucial points is very sparse, no doubt because of their sensitivity. Thus, even if the most is made of the evidence that is available, conclusions cannot be drawn with any degree of confidence.

2 What participation in the FBS involves

An interviewer visits each of the 20-25 families for whose budget data

she is responsible at least twice a month. For example, she may conduct her first series of interviews between the 1st and the 12th of the month, and her second series between the 16th and the 27th (Instruktsiya... 1960). Postnikov (1952) presents a typical interview schedule, reproduced by DePauw (1965 p.14). It allows for two three-hour interviews a day at various times between 8am and 10 pm, presumably depending on respondents' workshifts, for all adult members of the family are expected to attend the interviews. If a member of the family is absent, the interviewer must return within 3-4 days to see that person and complete the records. Thus interviews take up at least six hours a month of families' non-working time.

Families must also spend considerable time and effort keeping auxiliary records of incomes, expenditures and stocks, the format of which is reproduced in Table B5.1 (for workers and employees). Finally, they are expected to attend periodically meetings of participants at the workplace and elsewhere.

The procedure officially recommended to interviewers for a visit is as follows (Instruktsiya... 1960). First, the auxiliary records are inspected; if any days are missing, the transactions for those days are elicited and the records filled in. For example, the interviewer may visit on the 12th but records have only been completed up to the 10th: the interviewer must then fill in for the 11th and the 12th. The interviewer next collects information about the ready cash held by each family member. Then she adds up the incomes and expenditures entered in the auxiliary records, grouping expenditures under standard items on a special form (see Chapter C1) in accordance with detailed definitions given in the "Dictionary of Receipts and Expenditures".

Now the interviewer proceeds to the main part of the visit - the questionnaire-conversation (opros-beseda). The basic budget form, Form No.

TABLE B5.1

FORMAT OF AUXILIARY RECORDS KEPT BY FAMILIES

I. Money received (denezhnyi prikhod)

Day of month	By whom, for what and from whom received	How much received rubles kopeks	For notes
1	2	3	4

II. Goods received without payment (bezdenezhnye postupleniya)

Day of month	What received (milk, potatoes, eggs, boots etc)	From where received*	Quantity (indicate units)	For notes
1	2	3	4	5

*from own cow or garden, as present, from enterprise, from relatives etc.

III. Food products remaining at end of month

Name of food products	Quantity

IV. Expenditures

Day of month	On what spent	Where bought** or who paid	Quantity and unit of measurement	How much spent R K	For notes
1	2	3	4	5	6

**in State trade, in cooperative trade, in canteen, in buffet, on collective farm market, from individual citizens.

Source: Kats Ya D, Ocherki statistiki truda, M 1960, pp 174-5

1, which is reproduced by DePauw (1965), is filled in and checked by the interviewer while sitting with the family whom she consults as necessary. No corrections or changes to the form without the knowledge and consent of the family are allowed. Each item of income and expenditure listed on Form No.1 is considered in turn, with reference to the auxiliary records and to documents (receipts, wage records from the workplace bookkeeping office etc.) where appropriate[1]. The interviewer enters basic monthly indicators each month into a Check Notebook, and compares the figures for consecutive months as a check. If she and the family cannot explain a significant month-to-month discrepancy, this is taken as a sign of an error to be searched for.

When the family are away from home, they are still expected to keep auxiliary records, and to send them directly to the statistical administration. The interviewer then goes over the records with them on their return, if necessary carrying over corrections into the data for the next month.

To gain an impression of the scale of the task which must be fulfilled during an interview, let us consider the size of Form No.1. The figures usually cited are "more than 2000" questions for the "Budget of Worker and Employee" and "more than 3000" questions for the "Budget of Collective Farmer". However, Kolpakov (1968) gives "about 1500", and Rimashevskaya (1968) "1500-2000" questions for the former, while Kolpakov gives "about 4000" questions for the latter. The discrepancies may be explained by differing methods of reckoning up numbers of questions. Such huge figures are compatible with the level of detail of the classifications used in the version of Form No.1 ("Budget of Worker and Employee") presented by DePauw (1965 pp.53-61), which has 253 lines, most with several columns[2].

Let us mention two tasks for which a particularly heavy record-keeping burden prevents the collection of reliable data: the monthly balances of food products and the complex accounts relating to the husbandry of private

agricultural plots.

In compiling a balance of food products for a period (month or year), the interviewer must reconcile for each product the sum of the quantity stored by the household at the beginning of the period and the quantity received (by purchase, as payment in kind or from own production) during the period with the sum of the quantity consumed during the period and the quantity stored at the end of the period. Weights are supposed to be recorded to the nearest 10 grammes, volumes (of milk, alcoholic drinks etc.) to the nearest centilitre (DePauw 1965 p.50).

Families often fail to weigh and measure their stocks, and this must be why most interviewers - as we are told by Kalinichenko (1963) with regard to Kirgizia - try to visit families at harvest time to help them weigh fruit and vegetables collected from their private plots. It is hard to believe that the instruction forbidding interviewers to assess quantities approximately by eye is always observed.

Yet the need for these balances is a moot point. The economist A.Kh.Karapetyan (1980 p.249) advocates that the collection of data on stocks of food products be discontinued, as relative year-to-year changes in stocks are so insignificant that quantities obtained may be equated with quantities used: stock information is needed only where definite time trends occur, as with money savings.

The data which the budget survey tries to collect on the husbandry of private plots is very detailed and difficult to handle. Thus, net income derived from a plot must take account not only of sales of produce, but also of the value of own consumption and expenditure on implements, seeds, fodder and the like. The difficulty of establishing such incomes is such that it is now recognised that there are no reliable statistics on them, and they are reckoned under standard regional rates per family (Rzhanitsyna 1977)[3].

Collective farmers are also supposed to keep records of their labour outlays, an especially burdensome chore inasmuch as a person often performs several different types of work in the course of a single day (Kalinichenko 1963). Evidently the interviewer often fails to get the family to record the hours spent on different activities, for we are told of alternative methods used to estimate labour outlays: by interviewing family members, or by starting from the quantities of fruit and vegetables collected.

Soviet economists sometimes refer to the strain which participation in the survey imposes on families and the deleterious effect this has on the data collected:

> Compulsory, painstaking and regular record-keeping, and systematic visiting and questioning by interviewers sometimes evoke a negative reaction (Rimashevskaya 1968).
>
> The frequent visits by interviewers excite nervousness in families. They start to find their duties burdensome, and so the quality of their records deteriorates (Levin 1974).
>
> Continuous observation becomes extremely burdensome psychologically for families... This naturally affects the quality of records (Karapetyan 1980 p.37) [4].

The effort required is sometimes minimised in speeches made at meetings of participants: a Leningrad worker who has been in the survey for eight years claims that no special labour is involved provided that transactions are recorded promptly (Iz opyta... 1955). A certain A.I.Malygin, on the other hand, in the survey since 1932, admits:

> It is a very serious and tedious business, but very honourable and important work (Sobranie... 1963).

Such speeches must of course be interpreted with a view to the officially supervised formality of Soviet public life. They may well be vetted, if not written, in advance by the organisers of the meeting so that they express what are considered to be desirable attitudes.

Great emphasis is placed on the need for good relationships between interviewers and families. The ideal relationship is a common theme at

staff meetings:

> The interviewer must be a counsellor to the family, cultured, polite and modest (a Deputy Head of TsSU, Vsesoyuznoe... 1954).

Such a relationship is characterised by tact, mutual trust, understanding, respect and sincerity (Podgornov 1957, Sobranie... 1963). A favourable attitude of the family to their interviewer is a condition of successful work, for the interviewer "must closely concern herself with the internal life of the family" (Podgornov 1957). The meaning of this is made more explicit in the instructions for interviewers, which state that the respect and trust of the family must be won so that they will reveal "all their receipts, whatever the source, and all their expenditures, whatever the object" (Instruktsiya... 1960). Respondents may fear that sensitive information - for example, details of income from unofficial sidelines or of excessive alcohol consumption - will not be treated as confidential and will come to the notice of superiors at work or of Party and other vigilantes (Uchebnoe... 1958 p.195). Such fears, whether justified or not, are understandable under Soviet conditions, where TsSU cannot be perceived as clearly independent of other authorities.

To promote a close relationship with families, interviewers of collective farmer families are urged to take an active part in the life of the collective farm at which they are located, as this will bolster their authority. It is considered desirable that they live as well as work at the collective farm, and Shkrebel' (1957) claims that as a rule they do so. There seems to be no anxiety that intimacy with families may undermine the objectivity of the interviewer.

One could interpret the exhortations to maintain good relationships between interviewers and families as expressions of concern about a less ideal reality. Thus it is known that much expenditure on alcoholic drink

is hidden from interviewers (Chapter D4). It is moreover an inescapable part of the interviewer's job to press restive families to keep more careful records. A Deputy Head of the TsSU Budget Survey Department recommends that, when an interviewer discovers on her visit that a family member has not recorded his expenditure for several days, she insist that he record them there and then while she waits before proceeding with the interview, instead of the usual practice of doing it for him (Vladykin 1955). The delinquent will thereby see for himself how much more time he loses in this way, and will gradually learn to keep records on time. Certain tensions seem to be inherent in the interviewer-family relationship.

3 Incentives and pressures used to encourage participation

Long periods of burdensome participation in the survey can be achieved only by action to "preserve the sample" (Soveshchanie... 1964) - that is, to counter the inclination of families to drop out of the survey:

> Experience shows that, with rare exceptions, the population is not prepared voluntarily to keep records of income and expenditure over a long period... Great efforts and definite expenditures are needed for moral and material action upon the families selected in order to maintain a stable sample (Karapetyan 1980 p.37).

In this Section we describe the incentives and pressures used to encourage continued participation and conscientious record-keeping.

A small financial reward is received by families for participating in the survey (Yezhov 1965 p.317), and "receipts for participation in the budget survey" is indeed one of the income items listed in Form No.1 (DePauw 1965 p.38). The decree of the Council of Ministers on the budget survey promulgated on 3 November 1951 sets the payment at an average rate of 12 (old) rubles per month, the rate to be differentiated by Republic and oblast' (Resheniya... 1968)[5]. This was only about two per cent of the average wage of workers and employees, which was 64 (new) rubles in 1950

(Trud... 1968 p.137). Rimashevskaya confirms that payment continues on a small scale, of the order of 5 rubles per month (Shenfield 1982b), which is about three per cent of the current average wage of 169 rubles per month in 1980 (Sbornik... 1982 p.88).

Though this payment is quite small, families whose participation is judged exemplary or "most active" receive additional money bonuses and prizes from enterprise and collective farm managements, as well as such benefits under their control as holidays in sanatoria and Houses of Rest (Shkrebel' 1957, Brainin 1964, Sidlyarenko 1964, Rovinskaya 1965, Klimenok 1974). "Valuable commemorative presents" may also be given (Konovalov 1965).

These bonuses are awarded on quite a large scale. Of about 200 families[6] in the survey in East Kazakhstan oblast' in 1964, 49 were presented with bonuses or presents at a meeting of participants (Konovalov 1965). At a meeting in Kirgizia, with about 1000 participating families, in 1974, 175 bonuses were presented (Altunina 1974). Thus a substantial proportion of families must receive bonuses at some time.

Other material benefits are associated with attendance at the meetings of participants. These are usually followed by a free film show or concert. The arrangements for one meeting in Odessa for participants from Odessa oblast' amounted to a short holiday: hotel accommodation, an excursion round the town, a film and a concert, rounded off by a three-hour boat trip on the Black Sea (Klebanov 1968).

Material incentives are not then insignificant. They must bias the sample upwards in respect of total real income, as the sample is selected on the basis of the basic wage, exclusive of bonuses and free benefits. The incentives in kind may also be substantial enough significantly to distort the pattern of expenditure of the sample - that is, to divert expenditure from concerts, films or holidays, available free, to other

items. Strangely enough, Soviet critics of the FBS fail to make these points.

Material incentives are supplemented by moral ones. An effort is made to make respondents feel that they are carrying out important State work, appreciated by the authorities and the public alike. The Head of Minsk oblast' statistical administration assured a meeting that

> public opinion regards families who participate in the budget survey with enormous respect and gratitude (Dedyulya 1970).

Various institutions - TsSU, Trade Unions, the executive committees of oblast' Soviets - award medals for conscientious participation. These come in two grades: the Otlichnik Sotsialisticheskogo Ucheta (Person Excelling in Socialist Record-Keeping) and the lesser Pochetnyi Gramot (Mark of Honour). The lower-grade medal is awarded on a large scale. Thus, at a meeting in Moscow in 1964 (Rovinskaya 1965), the Head of TsSU in person, V.N.Starovskii, handed out 115 lower-grade medals, as well as an unspecified number of higher-grade medals.

A final benefit of participation is that families may be able to get help from their interviewers on matters unrelated to the survey. In effect the interviewer functions also as their social worker. As the Inspector of State Statistics for Pechora raion in the Komi ASSR (Filippov 1961) writes:

> Collective farmers sometimes turn to our staff with various requests, which we do our best to fulfil.

Shkrebel' (1957) gives checking interest on State loans and filling in pension forms as examples of jobs on which families ask for and receive advice and help. For many families, especially the poorly educated in rural areas, it must seem natural to turn to the interviewer, as the most accessible educated person, when confused by bureaucratic procedures.

Interviewers are officially urged to show willingness to provide such assistance, in order to promote friendly and cooperative relationships with families.

There is a little evidence that such assistance may extend beyond form-filling. Low-status families probably see in their interviewer a representative of "the authorities" who should be in a position to help them in all sorts of official dealings, from applications to enter educational institutions to trouble with residence permits.

Konovalov (1965) gives an unintentionally revealing account of how the interviewer can serve as a channel of influence (blat) for families who, were they not in the survey, would be virtually without influence in a society where much use is made of personal connections. The little daughter of a collective farmer family in East Kazakhstan fell ill and needed immediate surgery which was beyond the capacity of the raion hospital. The family mentioned their predicament to their interviewer, who spoke to her superior, the Head of the oblast' statistical administration who was also a Deputy to the oblast' Soviet. He in his turn arranged the girl's admission to the oblast' hospital with its Chief Doctor. The story has, of course, a happy ending: the little girl returns home after a successful operation. The right of Soviet citizens in general to medical treatment appears in a less happy light.

Material incentives to participate are backed up by a propaganda effort, going under the name of "mass-explanatory work", directed at participating families, and by close supervision of their participation. Interviewers are instructed systematically to inform the managements and "social organisations" (that is, Party and trade-union organisations) of the enterprises and establishments at which they are located "which families are successfully keeping records" and "what help other families need to improve their record-keeping" (Instruktsiya... 1960). This help is provided

through "Commissions to Assist the Budget Survey", formed at enterprises by order of oblast' statistical administrations and including management, Party and trade-union representatives (Brainin 1964). Commissions may even be created at the workshop level (Krasnoshchekov 1962, Lukasheva 1964). They "explain the significance of the survey" to the participants, answer their queries and organise meetings for them. Commission members are urged to get to know the families personally. Their status must enable them to exert considerable pressure on participants.

The mass-explanatory work conducted by and on behalf of the statistical offices uses the full range of available communications media: personal contact, workplace wall-newspapers, special magazines featuring photographs of the "most active" participants, the oblast' press, and occasional radio and television programmes (Vsesoyuznoe... 1954, Podgornov 1957, Konovalov 1965). Articles and programmes often focus on particular families, and publicity is given to the names of conscientious participants.

The propaganda medium to which most attention seems to be devoted is meetings of participants, both at workplace level and at city or oblast' level. The meetings are organised by officials of the statistical office, but enterprise functionaries and representatives of local leading bodies also take part. We shall describe a typical meeting at oblast' level.

First, preparations for the meetings must be made. Accommodation for the participants and the renting of a meeting hall must be arranged through the oblast' Soviet. Transport to and from the meeting may be arranged by the Soviet, or enterprise and collective farm managements may be prevailed upon to provide it. Interviewers make special visits to families to remind them to attend, invitation tickets may be issued, and managements are asked to rearrange participants' shifts as necessary. A crèche may be provided at the meeting[7].

The meeting is opened by a prominent TsSU official. Thus, at a meeting

in Moscow oblast' the Head of TsSU USSR L.Volodarskii

> thanked the workers and employees keeping budget records for their modest but exceptionally important State work, for their help to the bodies of State statistics in the analysis and study of questions of the growth in the welfare and culture of the Soviet people, and expressed his confidence that the workers and employees of Moscow oblast' would continue to keep budget records successfully and accurately (Sobranie... 1983).

There are then one or two lectures - one by a member of the statistical staff on "the significance of the budget survey", and perhaps one on a more general topic, such as "the rising standard of living in our oblast'" or the international situation. Speeches from the floor follow, mainly by survey participants who expatiate on what a rewarding experience participation is and urge one another on to yet greater efforts. A resolution on the need for more conscientious record-keeping may be put forward and unanimously passed. Bonuses and medals are presented. Greetings telegrams from top TsSU officials may be read out. The "business" part of the meeting completed, there is usually some kind of entertainment.

4 The refusal and drop-out rates

International experience leads us to expect that a large proportion of the households initially selected for a budget survey will refuse to participate in it or will drop out in the course of it:

> When the intensity of the survey and the intended detailed probe into all sources of income are explained to the prospective respondents, there is likely to be a high refusal rate, unless great pains are taken to obtain cooperation. Fatigue or annoyance, owing to the repeated visits by the enumerator that are necessary to complete the survey, may cause "drop-outs" during its progress (Casley and Lury 1981 pp.193-4).

It is assumed that an increase in the burden entailed by participation has the effect of reducing the response rate. The President of the Royal Statistical Society has expressed concern that the merger of the FES with

the National Food Survey proposed in the Rayner Report on the Government Statistical Services will reduce the response rate and hence the quality of the expenditure figures produced (News and Notes 1981).

Of about 5000 addresses selected annually for the British Family Expenditure Survey (FES) between 1957 and 1966, only 3000-3500 families cooperated, a response rate of 60-70 per cent (Kemsley 1968); by 1975 the response rate had stabilised at about 70 per cent. The response rate in the French survey was about 75 per cent of those contacted, themselves only about 80 per cent of those whose addresses were selected, in 1971 (Bigata 1973) and in 1978-9 (Glaude 1982). In some countries response rates are even lower - 30-40 per cent in Canada (Statistics Canada 1977), for example[8].

Experimental surveys in France have demonstrated a strong relationship between response rate and period of participation (Glaude 1982). The results are shown in Table B5.2. Almost all refusals to cooperate came on the initial approach; very few of those who then agreed to take part dropped out later. However, an attempt to use a twelve-month period of participation in Holland resulted not only in a response rate of 20 per cent but also in a quarter of those who began keeping records dropping out in the course of the year.

Taking the effective sample size as given, a low response rate harms the representativeness of results to the extent that respondents differ systematically from non-respondents. Studies of the differential response of different population groups do indicate that non-response is responsible for important biases[9]. Thus, comparison of the 1971 FES sample with corresponding census data revealed a pronounced decline of response rate with age, while rural areas had a much higher response rate (71-77 per cent) than London (58-64 per cent) (Kemsley 1975).

As participants in the Soviet budget survey bear a heavy burden over

TABLE B5.2

RELATIONSHIP BETWEEN RESPONSE RATE AND PERIOD OF PARTICIPATION IN EXPERIMENTAL FRENCH BUDGET SURVEYS

Period of participation(1)	Response rate per cent
Seven days	80
Ten days	75
Fourteen days	70
One month	below 50
Two months	below 50
One year	25

Note (1) That is, the period over which budget records were to be kept

Source: Glaude (1982)

very prolonged periods and have other reasons, such as fears concerning confidentiality, for being reluctant to cooperate, one would expect on the basis of international experience a very low response rate and a high drop-out rate. Soviet writers, however, assert the contrary. Matyukha, Postnikov and Samoilov (1958) state that "cases of refusal are very rare". An international meeting on the methodology of household surveys, convened in 1978 by the Conference of European Statisticians of the European Economic Commission of the UN, was induced to recommend "that Soviet experience of getting almost 100 per cent response be studied" (Dumnov and Riik 1978). Rimashevskaya also claimed in conversation with this writer that both non-response and drop-out rates are very low (Shenfield 1982b)[10].

Although some Soviet accounts of the FBS do refer to the occurrence of refusals to participate, without giving any indication of their frequency (Krylov 1957, Posobie... 1980), other accounts make no such reference (Matyukha 1966, 1967). Even the official instructions for FBS interviewers issued by TsSU in 1960 do not deal with the question of refusals (Instruktsiya... 1960). Similarly, Kildishev et al (1980) is one of very few sources which mention the need to replace "families who fail to keep records" by others with approximately the same characteristics. The 1960 instructions for interviewers do, however, cover this contingency.

One could in principle resolve the inconsistency between international experience with regard to refusal and drop-out rates and official Soviet claims on the matter in two ways: by dismissing the experience of surveys in other countries as irrelevant to Soviet conditions, or by dismissing the Soviet claims as false. The two alternatives do not exclude one another, and there is a certain amount of evidence in favour of each of them.

There are a very small number of Soviet sources which appear to conflict with the official claims. For example, Krutikov et al (1981 p.181) may be alluding to the frequent occurrence of refusals when they argue that

the "intellectual level" of participating families is somewhat above average because "families without the necessary inclination to keep records are automatically excluded". Again, Vagradyan (1962) gives as one of the advantages of surveys using the "method of momentary observations" (see Chapter B11) the achievement of a non-response rate of one per cent or less, implying that the non-response rate in the existing budget survey is significantly higher than this. Such hints, however, scarcely constitute conclusive evidence.

We are then led to question the relevance of international experience in this area. That Western and Soviet surveys are conducted in different social atmospheres is indicated by the very high response rates generally achieved in Soviet interview surveys. For example, the demographer Kiseleva reported that in a survey of women in Novgorod oblast' and the Chuvash ASSR on such a sensitive topic as fertility only one woman out of 2500 approached refused to be interviewed (Vsesoyuznoe... 1969 p.299). Professor Yu.V.Arutyunyan, a sociologist who specialises in rural problems, told this writer that Soviet sociologists are only now just starting to come up against the problem of refusals, and that most people still look on the chance to take part in a social survey as an interesting novelty (Shenfield 1981). This is perhaps of limited pertinence to the question of participation in the FBS, which can hardly stay an interesting novelty for very long.

Soviet writers put great emphasis on the principle of voluntary participation in the FBS:

> It is necessary to ensure in the most rigorous fashion that the workers selected have given their voluntary agreement to enter the survey. It is necessary to remember that... the voluntary principle is the basic condition of successful work... (Kats 1960 p.173; emphasis in original).

However, this emphasis seems to imply a real danger that in practice the voluntary principle may not be properly observed. We conjecture that

people are recruited to and kept in the survey, often with reluctance, by persuasion of statistical and workplace officials which amounts to pressure. This might explain the reference by Karapetyan (1980 p.253) to "persons specially attracted" to the work of keeping budget records. Persuasion to take part in such "important service to the State" may be convincing not only because it comes from people representing the authorities and who (in the case of enterprise officials) have personal power over the potential recruit, but also because of the possible ideological interpretation of refusal to cooperate:

> Soviet people readily supply the required information. There is no room for suspicion, caution and even hostility towards statistical surveys, which are typical reactions in capitalist countries... Soviet people know that the Soviet State conducts statistical surveys in their own interests (Yezhov 1967).

There is just one account in the literature of non-response posing a problem. A Deputy Head of the TsSU Budget Survey Department mentioned obstacles encountered when the sample was expanded in 1969 at a top-level staff meeting:

> At the end of 1968 Republican TsSU's selected enterprises, organisations, workers and employees for the additional sample. The families selected were sufficiently representative. However, here and there they did not trouble to obtain in good time the agreement of the families. When it came to recording incomes and expenditures, in some cases interviewers came up against families refusing to keep records. Replacements had to be found in a hurry, and sometimes they differed from the original families, harming representativeness (Sukhoruchkina 1970).

While this does show that the agreement of families cannot be automatically assumed and that non-response <u>can</u> be a problem, it also implies that agreement usually <u>is</u> obtained when sufficient time is allowed for persuasion[11].

It may be that families are persuaded to enter the survey, attracted by the prospect of payment for participation and not realising the effort

which will be involved. When they do realise this, any inclination to drop out is impeded by the pressure exerted to "preserve the sample", pressure suggested by a speech of P.I.Safonov, Secretary of the Party organisation at the Komintern factory, at a meeting of participants in Leningrad in 1955:

> Our workers have a serious attitude towards this important work. In the three years during which budgets have been collected at the factory, families have been replaced only in individual cases in connection with such circumstances as retirement, illness etc... (Iz opyta... 1955).

On the other hand, the operation of pressures characteristic of the Soviet system may not suffice to prevent substantial refusal and drop-out rates. It is also plausible to conjecture that the effectiveness of such pressures varies over time in response to changes in the social atmosphere. According to Syikova (1979), the refusal rate in the Bulgarian budget survey has sharply risen over time:

> In recent years an increasing proportion of families have refused to cooperate because of the long period of observation. Refusals occur both at the time of selection and after a certain time. In 1961 15 per cent of selected families refused to participate, but in 1972-4 the rate exceeded 32 per cent. This affects the organisation of the survey and the quality of the data.

In view of the basic similarity of the social system in Bulgaria and in the USSR, it is conceivable that the refusal and drop-out rates in the budget surveys of the two countries are broadly of the same magnitude.

If there is a significant refusal and drop-out rate in the FBS, then there is almost certainly a significant non-response bias. Even if, as TsSU instructions require, families who refuse to participate or drop out of the survey are always replaced by other families with approximately the same income, family composition and so on (Chapter B1), those who agree to participate are likely to differ systematically in other respects from

those who refuse. Those who agree to participate over long periods include a high proportion of the type of people generally inclined to conform to official expectations. Thus, a Trade Union official at a meeting of FBS participants in Moscow oblast' states that "many members of families keeping budget records for more than 10-15 years are advanced producers (peredoviki proizvodstva) who have been awarded orders and medals" (Sobranie... 1983)[12]. The budgetary behaviour of such people may well tend to diverge from that of their fellow workers.

5 Strategies of burden reduction

There are two strategies which families are thought consciously or unconsciously to use in attempts to reduce the burden of participation in the survey. One is negligence - not bothering to keep records completely, accurately or promptly. The other we call "regularisation" - the adjustment of budgetary behaviour itself in a way that makes it simpler to keep records of it. Negligence distorts survey data directly while regularisation makes sample budgets unrepresentative of the population as a whole.

We have already mentioned some types of negligence - in recording food stocks and private plot husbandry, and in failing to keep auxiliary records up-to-date. Here we shall discuss failures to keep auxiliary records continuously, and then the tendency to omit certain kinds of item from them.

In past decades a large proportion of participants did not keep auxiliary records because of inadequate literacy, so that reliance had to be placed on interviewing alone[13]. Although the keeping of auxiliary records has now long been established procedure, it may still not be universal practice; Berzkaln (1968b) writes that "the majority" (not all) "of surveyed families keep auxiliary records". But even if auxiliary records are never wholly absent nowadays, we have seen that they may be written up several days after the event, frustrating their purpose of avoiding reliance

on memory. As a result the interviewer may be told, for instance, that 10 kg of meat products were bought in the month when the true amount was 11 kg (Barbashov et al 1968 p.130).

The experience of Western budget surveys shows that one cannot rely on the respondent's memory[14]. Memory is less unreliable in the USSR, as Ofer and Pickersgill (1978) argue in defending their own retrospective budget survey of Soviet emigrants: incomes and prices (in State shops) are fairly stable, and economic matters are very high on citizens' minds. Nevertheless it is unsatisfactory to rely on memory for much of the information collected in the budget survey, including less important and less regular expenditures. When a respondent finds it hard to remember something and hesitates, the hard-pressed interviewer is tempted to suggest a plausible "standard reply":

> After 8-10 interviews the interviewer has "norms" in her head which - often unintentionally - she communicates to the collective farmer. For example, he has difficulty in establishing how many hours he spent on his livestock. The interviewer suggests, "An hour and a half?", and he willingly agrees (Uchebnoe... 1958 p.210).

Krastin' and Berzkaln (1972) report that auxiliary records "are kept unsatisfactorily in the majority of cases". Apart from deliberate concealment of certain kinds of transaction, such as expenditure on alcohol, tobacco and cosmetics, there is a tendency to neglect to record "petty" expenditures on such things as matches (Barbashov 1968, Rimashevskaya 1968), and also "petty" incomes (Vsesoyuznoe... 1959 p.273)[15]. Rimashevskaya (1968) adds that:

> The necessity of balancing income and expenditure is fraught with a certain distortion of data by families.

That is, the balance of income and expenditure, claimed by official accounts to guarantee accurate data, encourages families to "adjust" their records to show the required balance.

If negligent record-keeping can be regarded as a short-term strategy for reducing the burden of participation at the cost of tension in the relationship of the family with the interviewer and other supervisors, regularisation of budgetary habits can be regarded as a more comfortable long-term strategy for simplifying the work of record-keeping. A number of Soviet economists have drawn attention to this "guinea-pig effect" by which participating families over time become more deliberately organised and regular in their behaviour than the general population (Rimashevskaya 1968, Levin 1974, Karapetyan 1980). Krutikov et al (1981 p.178) attribute it to "the influence of heightened attention to their budget, a rational approach to housekeeping and an attempt to restrict expenditure on goods the consumption of which may be negatively evaluated" by the interviewer, such as alcohol and cigarettes. "In sociology this is called 'spoiling' of the object of observation, which changes under the influence of the observation itself" (Karapetyan 1980 p.36).

Regularisation remains no more than a plausible hypothesis in the curious absence of research into the question, either in the USSR or in the West[15]. But it is a process reflected in speeches at meetings of participants:

> The experience of our family convinces me that budget records allow the family budget to be well regulated, expenditures to be made more rationally (Sobranie... 1963).
>
> Budget records help daily life to be better organised and have important educational significance. They discipline and teach thrift. We have become accustomed to meet together as a family to discuss our budget (Rovinskaya 1965).
>
> The budget records which I have kept for fifteen years help our family to plan expenditures correctly (Panina 1974).
>
> Interviewer T.Kolokolnikova... emphasised that keeping a record of the family budget is... not only necessary to the State but also useful to the family itself, which is given the opportunity of seeing an exact picture of the distribution of its money outlays and incomes...

> Laboratory assistant... N.Volkova shared her twenty-years experience of keeping budget records. She said that correct planning of expenditure with the help of budget records greatly helps her family in housekeeping and in bringing up the children... (Sobranie... 1983).

These are clearly approved sentiments. It appears that TsSU officials regard the prevalence in the sample of behaviour which must be quite rare in the wider population not as a shortcoming but as a beneficial educational side-effect of the survey, indeed as a matter for pride. They thereby demonstrate a lack of understanding of the whole point of sampling as a method[17].

The biases generated by regularisation are a matter of speculation. The evident concern of regularised families with thrift is likely to entail biases against frivolous expenditures (including those which "may be negatively evaluated") and luxuries and towards savings. Rogova and Rakovskaya (1981) find discrepancies between FBS data on family savings and corresponding data from other sources of monetary statistics, and attribute them to regularisation. FBS data on year-average per-capita stocks of ready money are significantly lower than the accepted estimates for the population as a whole, while budget survey data on banked savings are more representative. Rogova and Rakovskaya explain this in terms of the "specially 'organised' turnover of money" within participating families; that is, they are less prone to leave unspent money lying around as ready cash.

6 Conclusions

Participation in the FBS is a heavy burden in time and effort, continuing over very long periods. Material and moral incentives, and propaganda and supervisory pressures, have to be applied to recruit those selected into the survey and to retain them in the sample thereafter. Families attempt to reduce the burden of participation by means of

strategies of negligence and regularisation of budgetary behaviour. In this situation both the quality of the data collected and representativeness of the sample are harmed.

Financial incentives - payment for participation, bonuses and prizes - bias the money income level of the sample upwards somewhat, while incentives in kind - presents, holidays and entertainment - bias the real income level of the sample upwards, and also bias the pattern of expenditure of sample families against the items used as incentives. Negligent record-keeping biases the apparent patterns of expenditure and income against "petty" expenditures and incomes. Regularisation biases the pattern of expenditure against luxury and frivolous items and against items which interviewers may be thought to disapprove of (alcoholic drink, tobacco, cosmetics etc.), and entails a bias towards banked savings (though not ready cash). All these biases are consequences of the social context within which the survey is conducted, involving prolonged participation under conditions of pressure.

In spite of the incentives and pressures used to encourage families to participate in the FBS, there may be substantial refusal and drop-out rates (although the evidence is inconclusive on this point) with consequent non-response bias.

Notes to Chapter B5

1 In one textbook the interviewer is recommended not to follow strictly the format of the survey form when questioning the family, but to conduct the interview flexibly as a free-flowing conversation pitched to the mental level of the people concerned. This is in contrast to Western surveys, in which the interviewer is usually instructed to follow precisely a standardised schedule of what to say and in what order, so that personal biases can be minimised. It is not clear whether this textbook reflects general practice (Uchebnoe... 1958 p.210).

2 However, in any particular interview many questions will not apply. The greater length of the form for collective farmers is mainly explained by the collection of data on the husbandry of private plots. The two variants of Form No.1 were unified in 1977 (Ekonomicheskaya... 1983 p.378). As there is no mention of the collection of any type of data being abandoned, the new form is probably about 3000 questions.

3 Estimates of food production and population incomes from private plots have been derived from budget survey data since the 1950s (Vsesoyuznoe... 1959, pp.23-4). They show the contribution of the private plots to collective farmer incomes falling from 48 per cent in 1940 to 32 per cent in 1970 and 25 per cent in 1980 (NKh SSSR v 1980 g, M 1981, p.385). Davydov (1981) argues that such figures under-state the role of the plots, and attributes this to faulty sample selection. The sampling indeed has faults, but it is doubtful whether the net effect of the faults we have identified - omission of families without able-bodied members and of families who share a cow with neighbours, undercoverage of regions to which access is difficult - is to bias the results in this direction. We are inclined to attribute any shortfall to incomplete records, in which distrust of the data collectors as well as the burden of record-keeping may play a part.

4 See also Berzkaln (1968) and Balansy... (1969).

5 One new (post-1961) ruble equals ten old rubles. Comparison with the allowances for interviewers' travelling expenses, also laid down in the decree, confirms that reference is to old rubles.

6 We are told that five enterprises, one State farm and two collective farms are covered. Assuming a standard 25 families at each of these eight places, we have 200 families in all.

7 The attention devoted to the conduct of meetings of participants varies. There is one complaint about Heads of oblast' statistical administrations and of their Budget Survey Departments not bothering to attend them (Soveshchanie... 1953).

8 The Israeli budget survey provides a good example of these problems. Conducted on quite a broad programme - incomes, expenditures, savings etc. - every four or five years with a sample of 2500 families, it used to have a response rate of 72 per cent, but this fell in the most recent survey to 60 per cent. As a result of differential response there were severe biases against such groups as large low-income households and one-person pensioner households. Ways of reducing the burden on the respondent are therefore being considered. (Sources: talk by Dr.M.Sicron, Scientific Director of the Central Bureau of Statistics of Israel, in June 1982, and conversation with CBS staff member M.Kantorovich.)

9 An attempt can be made to correct for non-response bias to a certain extent by reweighting the data, as is done in the French survey (Bigata 1973).

10 In accounting for this state of affairs Rimashevskaya argued that participation in the survey is "not such a great burden", thereby contradicting her own statements in Soviet sources. She also made much of the incentive effect of the small payment for participation. The reliability of her claim must therefore be doubted.

11 Zhutovskaya (1966) reports that in the 1958 microcensus there were almost no refusals, as a result of effective explanatory work and assurances regarding confidentiality.

12 Gidwitz (1982) describes peredoviki as small groups of favoured "front-rank workers", mostly Party members, who use advanced, more efficient equipment and receive large pay bonuses.

13 Thus Shcherbakov (1951) reports an increase in the proportion of collective farmer participants in three raiony of the Tatar ASSR keeping records from less than half to 75-78 per cent.

14 Consumption data given from memory at interviews are distorted in two opposite ways: forgetfulness leads to underestimated expenditure, while the "telescoping effect" - remembering events as being more recent than they really are - leads to overestimated expenditures for a given past period. The former distortion is stronger for minor expenditures, the latter for major expenditures. Research in France in 1971-2 showed that respondents, asked what clothing they had purchased in the last month, overestimated their expenditure through telescoping by 61 per cent (Glaude 1982).

15 Incomplete auxiliary records is a serious problem in Western budget surveys also. In a French experimental survey respondents kept records for ten days, with the assistance of the interviewer on the first day only. Average recorded expenditure on the third and tenth days was only 90 per cent and 85 per cent respectively of that on the first day (Glaude 1982). Sharot (1982) found that expenditures recorded in diaries by housewives on a long-term British consumer panel for a market research firm (Audits of Great Britain) were too low on average by 20 per cent.

16 Research can fail to confirm even the most plausible hypotheses. For example, an experimental survey in West Germany failed to support the idea that people who volunteer to take part in a publicised budget survey have consumption patterns unrepresentative of the population as a whole (Glaude 1982).

17 It appears that it is not only in the USSR that official statisticians take a benevolent attitude towards the guinea-pig effect. In an article on the plans for the West German budget survey of 1973, Euler (1972) states that volunteers are expected to come forward motivated by civic duty and the prospect of a small money payment. It will be explained to people that participation will be useful to them personally as well as to the community: through systematic record-keeping they will gain a clear picture of their budget.

CHAPTER B6

CHECKS OF SAMPLE REPRESENTATIVENESS

1 Introduction

Official accounts of the FBS describe procedures which supposedly check sample representativeness at each stage of sample selection and provide for the correction of significant biases thereby discovered. This chapter is devoted to an examination of these checks. The basic principles of representativeness checks, and the special problems of applying them to the budget survey, are explained in Section 2. In Section 3 we describe and criticise the representativeness checks carried out by TsSU.

The calculation of estimated sampling errors should play an important role in representativeness checks. Although it does not play such a role in Soviet practice, some research work in this area has been carried out, and is reported and assessed in Section 4. In Section 5 we summarise our conclusions and discuss possible reasons for the deficiencies of TsSU methodology.

In this chapter we consider only the "standard" representativeness checks, regularly conducted by TsSU, in which sample data are compared with corresponding data for the restricted territorial and branch subpopulations covered by the survey. In view of the far from complete coverage of the Soviet population by the survey, these checks cannot assess the extent to which the sample is representative of the population of the USSR, or of regional subdivisions of the USSR, as a whole. There are other "non-standard" representativeness checks which do permit such an assessment, and these are discussed elsewhere in the thesis.

The "non-standard" checks are of two types. First, checks have been made, both by TsSU and by research economists in outside institutes, of the distribution of the sample by per-capita income, by comparison with population distributions by per-capita income derived from other sources. These checks, and their use for the derivation of corrective coefficients for adjusting budget data, are dealt with in Chapter B9. Second, checks have often been made by economists of budget data on consumption against corresponding regional and All-Union sales figures taken from retail trade statistics. Figures from both types of "non-standard" check are used in our general assessment of sample representativeness in Chapter B7.

2 The principles of representativeness checks

The closer that estimates of population characteristics obtained from a sample come to the, generally unknown, true values of the population characteristics, the more representative the sample is judged to be. The inaccuracy of a sample estimate - that is, its deviation from the true value - is the sum of two components, the sampling error of the estimate and its bias. Any particular sample actually selected is a member of the hypothetical population of all the samples which could have been selected, given the sampling design used; any sample estimate actually obtained is a member of the population of all the sample estimates which could have been obtained. Its sampling error is its deviation from the mean of this population of possible estimates, while its bias is the deviation of this mean from the true value. Thus, sampling error is the product of imprecision inevitably associated with the use of sampling. Its probable magnitude can be estimated theoretically from the size and design of the sample, which should be determined with a view to keeping sampling error within acceptable limits. Bias comprises systematic error which remains when sampling error has been accounted for. It is an aim of sample design

to prevent any significant bias, the presence of which indicates some failing in sample selection.

Provision should be made in a sample survey for representativeness checks to ascertain whether the sample actually obtained is as representative as planned and, if not, to make it possible to apply approximate corrections to the data and to improve the sample design in future surveys. Sample data are deliberately collected for which corresponding population statistics are already available as standards for comparison - for example, geographical, occupational and age breakdowns from a recent population census. The representativeness checks are passed if differences between sample and population data can, with a high degree of probability, be attributed to sampling error as theoretically estimated. If the differences are too large or too systematic in direction (or both) to be thus attributed, the pattern made by them is studied in order to judge whether sampling error is for some reason greater than planned and what biases, if any, seem to be present.

As Moser and Kalton (1971 p.268) point out, the failure of such checks reveals ways in which a sample is unrepresentative, but their success can never conclusively prove that a sample is fully representative. A sample which proves satisfactory with regard to all the variables for which population data are available may still be unrepresentative with regard to other variables for which population data are not available, including the most important variables for the study of which the survey is conducted. If representativeness checks were possible for the main survey variables, there would be no need for the survey.

However,the value of representativeness checks can be enhanced if they can be carried out on variables which are known to be highly correlated with the main survey variables. The Laboratory of Sampling Methods of NII TsSU have demonstrated this by means of experimental calculations using

data from the budgets of 418 families of workers and employees in an unnamed Republic (Venetskii and Chernysheva 1978). They concluded that, if a sample is known to be representative with regard to variable A, it can be taken as representative with regard also to any variable B the correlation coefficient of which with variable A is at least 0.4. Thus, if it is verified that a budget sample is representative with regard to per-capita income, family size, wage of selected person and per-capita family wage, then the sample can be assumed representative with regard to per-capita expenditure on food goods, non-food goods, services, footwear, meat, butter, public dining and knitted goods. This indirect method cannot, of course, be applied to any survey variable which is not highly correlated with the feasible check variables.

Introduction of the procedure advocated by Venetskii and Chernysheva depends on the availability of reliable population data on the four check variables listed by them. Family size is given by the population census, and there are ample statistical report data on wages. Use of the three-yearly incomes survey (microcensus) as a source of population data on family incomes is proposed, but the sample of this survey is itself subject to severe bias (Chapter B1) and is scarcely suitable as a standard against which to check other data. Proper representativeness checks for the budget survey require improvements in other areas of Soviet statistics.

The potential value of representativeness checks in the family budget survey is limited by the fact that the survey covers only certain sections of the population (for example, skilled and semi-skilled workers in branch X in Republic Y). Representativeness checks can be and are carried out for each such section of the population which is covered (Section 3). These checks, even were they competently conducted, could only verify the representativeness of partial samples of little substantive interest. Checks of sample data against data relating to the population as a whole,

on the other hand, always show what was anyway already known - that is, the extreme unrepresentativeness of the sample at this level and the need for reconstructing it on other principles. The theoretical rationale of representativeness checks properly applies only when a survey is designed to be fairly representative in the first place.

3 The standard representativeness checks

The standard procedure for checking the representativeness of a sample of workers[1] selected in an oblast' is described by Matyukha (1966). For each enterprise selected, the average monthly wage of the workers selected is compared with the average wage of all the workers in the enterprise (presumably excluding the categories of workers not covered by the survey). Further checks, for each branch covered and for all the branches covered taken together, are illustrated in Table B6.1. The average wage of each sample of workers (col.5) is compared with the average wage of all workers in the enterprises covered (col.7) and with the average wage of the corresponding population of workers in all enterprises of the branch(es) (col.8). If there is a large discrepancy between sample (col.5) and population (col.8), it is possible to trace at which stage of the selection process it has mainly arisen: the selection of enterprises is assessed by comparing columns 7 and 8, the selection of workers within enterprises by comparing columns 5 and 7[2]. A comparison is considered to indicate a satisfactory level of representativeness if the sample average deviates from the population average by not more than $\pm$ 5 per cent[3].

As we noted in Chapter B1, the wage of the selected individual is a poor indicator of the level of family income, and its use as sole check variable is often criticised as inadequate (Rimashevskaya 1965, Shvyrkov 1965, Karapetyan, Rimashevskaya and Sidlyarenko 1967, Korovkin 1969). However, the lack of family-related population data by branch obstructs

TABLE B6.1

REPRESENTATIVENESS CHECKS I: AVERAGE WAGES OF THE WORKERS SURVEYED IN AN OBLAST'

Branch of industry surveyed in the oblast'	No. of workers selected in the branch	No. of other family members in the branch(1)	Total no. of workers in the branch in selected families(2)	Average monthly wages (rubles) of: workers selected in the branch	workers in the branch in selected families(2)	all workers of the enterprises selected in the branch	all work of all enterpri in the branch
Col. 1	Col. 2	Col. 3	Col. 4 = Col. 2 + + Col. 3	Col. 5	Col. 6	Col. 7	Col. 8
Engineering and metalworking	180	25	205	159	148	152	155
Coal-mining	100	15	115	208	190	200	188
Textiles	50	100	150	130	125	132	129
All three branches surveyed in the oblast' taken together	330	140	470	158	175	165	159

Notes: (1) That is, those other members of the families of workers selected in the branch who also work in the same branch.

(2) That is, all members of the families of workers selected in the branch (including the workers selected) who work in that branch.

Source: Matyukha (1966 p 82); figures hypothetical

the use of the additional check variables (family composition, per-capita income etc.) proposed by critics[4]. The closest that the representativeness checks come to the use of family income as a check variable is consideration of the wages of other members of the families of selected workers who work in the same branch (cols. 3,4,6)[5], not an obviously worthwhile exercise.

It is sometimes stated that checks of average wages are, or should be, supplemented by comparison of sample and population wage distributions, using data from the periodic censuses of wage distribution by branch (Matyukha 1966, 1967; Soveshchanie.... 1969). An example of such a check is given in Table B6.2; Matyukha considers the sample distributions for coal and textiles insufficiently representative, but gives no criterion for making this judgement. However, "in practice representativeness checks are confined to the comparison of average values" (Venetskii and Chernysheva 1978). This shortcoming is especially serious in view of the fact that the sampling procedure tends to generate samples with smaller wage variability than the population. This is a consequence not only of the "tail-cutting bias" (Chapter B1), exacerbated by the asymmetric population distribution, but also of the practice of correcting sample averages which deviate too far from population averages by replacing extreme sample units by new units close to the average. Venetskii and Chernysheva (1978) attempt to convince TsSU officials of the need for checks of distributions, and to explain the use of the chi-square test as a criterion of correspondence of sample and population distributions, but the effect of their article on practice is unknown. Change requires that a long-established concept of "representativeness" as determined by averages alone be superseded; a "representative sample" is still defined as one the average indicators of which do not to an impermissible extent deviate from the corresponding average population indicators (United Nations 1978).

The assessment of discrepancies between sample and population averages

TABLE B6.2

REPRESENTATIVENESS CHECKS II: WAGES DISTRIBUTION OF THE WORKERS SURVEYED IN AN OBLAST'

Classes of workers by average monthly wage	Engineering and metalworking branch		Coal-mining branch		Textiles branch		All three survey branches taken together	
	All workers in selected families (1)	All workers in the branch	All workers in selected families (1)	All workers in the branch	All workers in selected families (1)	All workers in the branch	All workers in selected families (2)	All workers in the three branch
rubles	%	%	%	%	%	%	%	%
0·0 - 40·0	0·1	0·1	0·0	0·0	0·2	0·2	0·1	0·1
40·1 - 50·0	1·5	1·6	0·1	0·2	2·3	2·5	1·8	1·5
50·1 - 60·0	3·5	3·4	0·9	1·0	5·5	6·0	3·9	3·4
60·1 - 70·0	3·8	4·0	2·4	6·5	11·0	7·0	4·8	6·5
70·1 - 80·0	5·0	4·8	3·1	3·0	16·8	16·9	8·5	8·5
80·1 - 90·0	9·5	10·1	6·2	6·5	20·0	20·2	12·5	12·0
90·1 - 100·0	13·8	13·5	10·9	10·6	18·1	18·8	12·5	12·6
100·1 - 120·0	15·6	15·7	12·3	12·3	12·4	16·3	15·6	16·0
120·1 - 140·0	20·0	20·1	13·2	13·0	8·0	8·0	18·0	19·0
140·1 - 160·0	15·0	15·1	18·2	18·3	4·9	4·9	14·0	14·1
160·1 - 200·0	8·0	7·9	18·0	14·0	0·8	0·2	6·0	4·0
200·1 - 300·0	3·0	2·9	10·8	10·5	0·0	0·0	2·0	2·0
300·1 -	1·2	0·8	3·9	4·1	0·0	0·0	0·3	0·3
	100·0	100·0	100·0	100·0	100·0	100·0	100·0	100·0

TABLE B6.2 (cont'd)

Notes (1) That is, all members of the families of workers selected in the branch (including the workers selected) who work in that branch.

(2) That is, all the workers covered by note (1) taken together.

Source: Matyukha (1966 p 83); figures hypothetical

in percentage, rather than absolute, terms is also unsound. Percentage deviations cannot readily be compared with sampling errors, which are calculated as absolute quantities. Let us show how misleading it is to use a criterion unrelated to population variability by means of a numerical example. Application of the $\pm$ 5 per cent rule to a population with an average wage of 120 rubles leads to a sample being considered sufficiently representative provided that its average wage falls in the range 114 - 126 rubles. This may be appropriate if there is a very wide range of wages within the population, say 60 - 200 rubles, but becomes an extremely lax criterion if the range is fairly narrow, say 100 - 140 rubles.

Representativeness checks of collective farmer samples follow principles similar to those of worker samples. The main difference is that checks are carried out on not one but quite a large number of variables - seventeen at the stage of selection of collective farms alone, according to Shvyrkov (1965). Indicators used are listed in Table B6.3. Multiple checks are made possible by the availability of population data in appropriate form from such sources as annual collective farm reports and the quarterly census of privately owned livestock (Sukhoruchkina 1970). The representativeness of the sample of collective farms selected in an *oblast'* is checked first, and then that of the sample of collective farmer households by comparison (a) with all households in the selected collective farms, and (b) with all households in all the collective farms in the *oblast'* (Kildishev *et al* 1980). A check of a collective farmer household sample is illustrated in Table B6.4; such a check is conducted for each collective farm taken separately as well as for all the selected collective farms taken together. Matyukha (1966) also mentions a check of income distribution of collective farms (Table B6.5), but there is again doubt regarding whether such a check is carried out in practice.

If the system of representativeness checks for the workers' survey is

TABLE B6.3

INDICATORS USED FOR REPRESENTATIVENESS CHECKS IN THE SURVEY OF COLLECTIVE FARMER FAMILIES(1)

A. INDICATORS PERTAINING TO COLLECTIVE FARMS - for representativeness checks of samples of collective farms in oblasti etc.

Population present

Number of children

Distance from the raion centre and from the nearest railway station

Number of labour-days worked

Number of labour-days worked per household

Payment per labour-day (in money, in products)

Total money income of collective farm

Total money income of collective farm per household

Number of able-bodied men and women

Total sown area(2)

Sown area of grain and beans, of potatoes, of vegetables, of technical crops(2)

Number of head of collective farm livestock(2)

Number of head of collective farm livestock by basic types (horned cattle, sheep and goats, pigs)(2)

B. INDICATORS PERTAINING TO COLLECTIVE FARMER HOUSEHOLDS - for representativeness checks of samples of collective farmer households in collective farms

Number of persons in household

Number of persons present

Number of able-bodied persons

Number of able-bodied persons working on the collective farm

Number of persons working in State and cooperative organisations

Number of labour-days worked

Money value of a labour-day

Total sown area of private plot

Area of private plot used for vegetables and melons, for potatoes

Total head of livestock in personal ownership

Head of livestock in personal ownership by type (all horned cattle, cows, sheep and goats, pigs)

cont'd...

TABLE B6.3 (cont'd)

Notes:

(1) Sources used were Postnikov (1953), Ananyeva (1964), Kozlov et al (1965), Matyukha (1966), Kildishev et al (1980) and Posobie... (1980 p.373). The lists of indicators given in different sources vary somewhat; all indicators shown in any of the sources are included in the Table. The Table may therefore be incomplete, but it may also include indicators rarely or no longer used. Shvyrkov (1965) states that 17 indicators in all are used in representativeness checks of samples of collective farms (list A).

(2) According to Matyukha (1966), but not according to other sources, these are calculated per household.

TABLE B6.4

REPRESENTATIVENESS CHECKS III: AVERAGE BASIC INDICATORS OF COLLECTIVE FARMER FAMILIES[1]

Indicator	Sample of collective farmer families	Population of collective farmer families	Deviation[2] of sample mean from population mean
			%
Mean number of persons per family	4·2	4·0	+ 5·0
Mean number of able-bodied persons working on the collective farm per family	2·2	2·1	+ 4·8
Mean number of person-days of work on the collective farm over one year per family	682	679	+ 0·4
Mean money equivalent of one person-day of work on the collective farm (rubles)	4·20	4·16	+ 1·0
Mean area of family private plot (hectares)	0·25	0·24	+ 4·2
Mean number of head of livestock owned per family:			
large horned cattle (including cows)	1·5	1·5	0
cows	1·0	1·0	0
pigs	2·1	2·1	0
sheep	3·2	3·3	- 3·0

Notes: (1) The comparison is carried out (a) for the sample and the population in each of the collective farms surveyed, and (b) for the global sample and the global population of collective farmer families in the oblast'.

(2) Given in the source in the form of ratios of sample means to population means (105·0 per cent etc).

Source: Matyukha (1966 p 88); figures hypothetical

TABLE B6.5

REPRESENTATIVENESS CHECKS IV: DISTRIBUTION OF COLLECTIVE FARMS SURVEYED IN AN OBLAST' BY ANNUAL LABOUR PAYMENT MADE PER COLLECTIVE FARMER

Classes of collective farms by amount of annual labour payment made per collective farmer	Distribution of the population of collective farms in the oblast'	Distribution of the sample of collective farms surveyed in the oblast'
rubles	%	%
0·0 - 60·0	1·2	1·1
60·1 - 96·0	2·8	3·0
96·1 - 132·0	3·3	10·8
132·1 - 180·0	6·0	5·8
180·1 - 240·0	8·8	8·7
240·1 - 300·0	10·9	10·9
300·1 - 360·0	14·2	14·0
360·1 - 420·0	16·5	16·4
420·1 - 480·0	21·0	21·1
480·1 - 540·0	10·1	3·2
540·1 - 600·0	3·0	2·9
600·1 -	2·2	2·1
	100·0	100·0

Source: Matyukha (1966 p 86); figures hypothetical

inadequate because only one check variable is used, that for the collective farmers' survey probably suffers from the use of too many check variables. When checks of a sample of collective farms in an oblast' show up "significant" discrepancies between sample and population averages, one is supposed to attempt to "correct" the sample by replacing individual collective farms by others of the same productive orientation, and desirably in the same raion, and then to repeat the checks (Kozlov et al 1965 p.230)[6]. It must be impracticable to continue such a process of trial and error until all seventeen variables pass their checks simultaneously unless a computer is used. Berzkaln (1968) describes a computer algorithm developed for this task by the Latvian Division of NII TsSU, but there is no indication of its use in practice. Therefore the theoretical procedure for correcting discrepancies cannot be fully implemented, and samples must remain biased on many of the check variables. Thus, Starovskii in 1970 urged that an effort be made to guarantee representative survey data on privately owned livestock, so that the censuses of privately owned livestock could be discontinued (Sukhoruchkina 1970), but this has not occurred.

There is little information on the frequency with which representativeness checks are conducted. It appears that in the RSFSR they are carried out annually on a set programme (Sukhoruchkina 1970)[7], but one also finds complaints that they are completely neglected in some areas - for example, Uzbekistan (Krasnoshchekov 1962).

Further limitations in the conduct of representativeness checks are revealed in the discussion by Kozlov et al (1965 pp.230-1) of "frequent errors" made by users of budget survey data. Users should not assume that data are representative on indicators not covered by the checks, or at territorial levels below those used in the checks. Collective farmer samples are checked by oblast' in the RSFSR and the Ukraine, but elsewhere (for example, Kazakhstan) only by Republic. Worker samples are checked

by branch for the USSR as a whole and for economic regions, each of which consists of several *oblasti* and/or small Republics; Matyukha's account of checks by *oblast'* is thus presumably misleading. Kozlov *et al* warn that, even within these limits, "representativeness checks show that the sample deviates from the population on basic budget indicators". Correction of discrepancies by replacing sample units can then be far from automatic.

The need for calculating theoretical sampling errors as part of the study of sample representativeness has long been recognised by various TsSU officials. The start of such work in the Ukraine is reported in Krasnoshchekov (1962), while sampling error bounds were calculated for the RSFSR in 1963 and 1966 (Sukhoruchkina 1970). Samoilov, Deputy Head of the Budget Survey Department of TsSU USSR, in 1970 attributed the insufficient use of sampling error limits to the laboriousness of manual calculations, and stated that from 1971 they would be calculated by computer "for a significant number of basic indicators".

In the event, however, computerisation has not been accompanied by any substantial methodological changes, and it is clear from Venetskii and Chernysheva (1978) that the traditional method of conducting representativeness checks remains in use[8]. A recent description of the "Budget Statistics" subsystem of the Automated System of State Statistics makes no mention of sampling error calculations (*Avtomatizirovannaya*... 1979 pp.288-9), and a lecture on the computer system delivered to a staff meeting in 1979 likewise makes no reference to them (*Soveshchanie*.... 1979). In an account of the computer subsystem by Ananyeva, Ivanova and Khlopina (1980), sampling error calculations are discussed, but still under the heading of "methodological problems" for future solution (See Chapter C4).

4 The calculation of sampling errors

Quite extensive calculations of sampling errors for the budget survey are reported by Shvyrkov (1965) and by Venetskii and Matyukha (1968). We briefly review their work, and then consider the nature of the fomulae they use.

Shvyrkov (1965 pp.28-36) explores the relationship, for a sample of given size, between the sampling error of consumption variables and the structure of the sample in terms of income, size and composition of families. For this purpose he calculates sampling errors for variously constructed subsamples of the worker and collective farmer samples for 1960 - subsamples heterogeneous in terms of family composition and income, subsamples homogeneous in terms of family composition but heterogeneous in terms of income, and subsamples homogeneous in terms of family composition and income.

An example of his findings is illustrated in Tables B6.6 and B6.7. As the average income of the population rises over time, so do the variances of consumption variables. Thus absolute sampling errors tend to rise, but more slowly than average consumption, so that relative sampling errors (sampling error/mean) tend to fall, as shown for workers in the period 1953-1960 in Table B6.6. However, in making territorial comparisons at a given point in time, the effect of a differential in average income may be outweighed by other factors. Thus, in Table B6.7, relative sampling errors are greater in oblast' "B" than in oblast' "A", although average family income is 12 per cent higher in oblast' "B". This is because "B" is a northern oblast', with a much more seasonal consumption pattern than the southern oblast' "A". The positive association between sampling error and the seasonality of consumption of food products is more fully dealt with by Filippova (1962).

Shvyrkov also categorises budget indicators according to their level

TABLE B6.6

SAMPLING ERRORS AT ALL-UNION LEVEL

Indicator	Twice the relative sampling error for the USSR sample of workers in	
	1953	1960
	%	%
Total family income	0·50	0·41
Total family wage	0·54	0·43
Family expenditure on:		
food	0·32	0·27
clothing and footwear	0·54	0·45

Source: Shvyrkov (1965)

TABLE B6.7

SAMPLING ERRORS AT OBLAST' LEVEL

Indicator	Twice the relative sampling error for the sample of collective farmers	
	in the southern oblast' "A" (1)	in the northern oblast' "B" (1)
	%	%
Number of persons in family	2·1	3·7
Proportion of family members who are children	4·5	6·5
Per-capita income of family	2·1	2·5
Family expenditure on:		
food	1·4	2·2
eggs	1·3	2·5
milk	1·8	2·6
dried fruit	1·4	3·9
clothing, linen, fabrics and footwear	2·2	2·8
furniture	3·8	3·1
cultural and everyday needs	2·1	3·0

Note (1) The source does not reveal the identity of the oblasti.

Source: Shvyrkov (1965)

of relative sampling error in comparison with that of family income. The most frequently and regularly occurring items, such as the payment of pensions and grants and the purchase of potatoes and bread, have the smallest sampling errors, and the least regularly occurring, such as the purchase of furniture and household goods, have the largest sampling errors.

Venetskii and Matyukha (1968) calculate the sampling errors of several average income and expenditure indicators (a) for 720 budgets of workers employed in six industrial branches in an oblast' "A", and (b) for 100 budgets of workers in the oil industry in an oblast' "B". They then estimate sampling errors for the whole USSR sample of 16,700 industrial workers, using for the purpose the sample variances obtained from the subsamples (a) and (b). These subsamples are so small and unlikely to be approximately representative that the exercise must be regarded as purely illustrative in nature. From the result that, for the USSR as a whole, relative sampling errors do not exceed 2-3 per cent, the unwarranted conclusion is drawn that "this means that the sample is fully representative, and its data may indubitably be used in practical work" - as if small sampling errors suffice to rule out bias, the other component of inaccuracy. There seems to be a certain tendency, here and elsewhere (as in Sukhoruchkina 1970), to regard sample-population comparisons and sampling error calculations as alternative rather than complementary methods of checking sample representativeness, an attitude consistent with the position of mathematical statistics as a special discipline separate from "general" statistics.

In the work reviewed above, the formulae used for calculating sampling errors are those appropriate for a stratified random sample. The budget survey sample, however, is more complex than a stratified random sample in two main ways: (1) sample units are not spread fairly evenly over the population of each stratum, but are clustered in workplace clusters; and (2) selection is not strictly random, but systematic with wage etc. as

the ordering variable (Chapter B1). The neglect of clustering in the formulae chosen leads to a significant underestimation of sampling error, and this is especially serious in view of the fact that the clusters are based on workplace, not area of residence. Intra-cluster correlation among the families of workers employed in the same enterprise is greater than that among families living in the same population point, and one of the arguments for reorganising the survey on the territorial principle is therefore that this would reduce sampling error (Ananyeva 1966, Safronova 1968; see Chapter B10). On the other hand, systematic sampling gives considerably smaller sampling errors than random sampling when the ordering variables are correlated, as here, with the variables under study, and the consequent overestimation of sampling errors must to a large extent cancel out the underestimation which results from ignoring clustering. The sampling errors calculated may then be at least of the right order of magnitude.

Nevertheless, the failure even to notice these points suggests a rather limited acquaintance with sampling theory on the part of the authors. Similar deficiencies are far from rare in Western survey work as well, though not in the work of the most professional organisations. In the USSR, however, these errors appear to be the rule rather than the exception.

Thus, Shlyapentokh (1976) observes that, although cluster sampling is very common in Soviet sociological surveys out of practical considerations, it is used without a theoretical understanding of the problems it entails and "the consequences are neglected". A striking example of this, apart from the budget survey, is provided by the sample survey of rural conditions in four <u>oblasti</u> conducted under the leadership of Arutyunyan (1971 pp. 340-368). Although both <u>raiony</u> and then villages are selected as sample clusters within each <u>oblast'</u>, the clustering is ignored in calculating the necessary sample sizes and sampling error. Part of the explanation is no doubt the inadequate, and sometimes actually incorrect, treatment of cluster sampling

in textbooks of mathematical statistics, of which Kazinets (1973) complains.

The use of systematic sampling in Soviet surveys is discussed, with special reference to the budget survey, by Druzhinin (1975), the mathematical statistician who has perhaps done the most to re-establish sampling theory in the post-Stalin period. He attributes the wide use of systematic sampling not to a theoretical understanding of its properties but to its practical simplicity and to "the intuition that systematic sampling is more precise than random sampling" in distributing the sample evenly through the distribution of the ordering variable. Druzhinin explains the conditions under which this intuition is correct, advocates the use of corrections to take account of the tails of the distribution omitted in systematic sampling with mid-interval starting-point, and considers how sampling errors can be estimated. This is a very difficult problem: systematic sampling is strictly speaking a type of stratified sampling in which the strata are defined by the intervals on the ordering variable, but the intra-stratum variances needed to calculate the sampling error of a stratified sample cannot be estimated, as only one unit is selected in each stratum. When there is a linear relationship between the variable under examination, Y, and the ordering variable, however, it is possible to make a very approximate estimate of sampling error on the basis of the sequential differences between the Y values in successive intervals; Druzhinin's account of the method makes reference to Western textbooks. His illustration of the application of the method to budget survey data does not convincingly show that the vast amount of computation required for its large-scale use would be worthwhile: the method gives a variance of sample wage averages more than 30 per cent smaller than the true value. Further research could perhaps arrive at some rough correction factors to take account of systematic sampling.

5 Concluding discussion

The system of representativeness checks which we have examined is technically deficient in many ways. The standard checks of branch-territorial subsamples against the corresponding subpopulations cannot by their very nature reveal the most serious biases in the sample, those arising from the total exclusion from coverage of large segments of the Soviet population. It seems that action is not often taken to correct even those discrepancies which the checks do reveal. The function served by the checks against retail trade statistics, which alone provide for comparison with data relating to the entire population of different territories, is not clear. Improvement of the methods used, and even research directed at such improvement, have proceeded so far at a painfully slow pace.

However, it would be wrong to conclude that the checks of sample representativeness are nothing but a pointless ritual. The work implies recognition that sample representativeness, however inadequately understood, <u>is</u> a goal to be striven for. The results of the checks, especially of those against trade statistics, must constantly draw the attention of official statisticians to the deficiencies of the existing sample, and provide critics of the survey with indisputable evidence in support of their complaints and proposals for change. In the absence of any study of the representativeness of the sample, there would be no basis for discussion of reform at all. Moreover, the results of checks do, at least from time to time, prompt action to improve sample representativeness (or at least to avert further deterioration). Some types of representativeness check have also played a part in the derivation of corrective coefficients, the application of which has made budget survey data somewhat more reliable and usable (Chapter B9).

We have seen how the established representativeness checks are in practice concerned only with the average values of variables, to the

exclusion of all other characteristics of their distributions. This is a natural corollary to the emphasis in the programme for processing survey data, to be considered in Chapters C3 and C4, on the generation of average quantities. Venetskii and Chernysheva (1978) press for checks of distributions on the grounds that such checks are a necessary precondition of the "differentiated approach" to the analysis of budget data, involving their breakdown by socio-economic groups[9] (defined on the basis of per-capita income, family composition etc.). However, TsSU has put up great resistance to the inclusion of such breakdowns in the data-processing programme (Chapter C3).

The inadequate methodology of the representativeness checks can best be accounted for in historical terms. As we noted in Chapter A4, it is common Soviet practice to rely on such checks in comparisons of population and sample means[10], and this practice reflects the approach taken to sampling at an earlier stage of its development. The survival of the practice appears therefore to be one of the results of the "freeze" imposed on the application of mathematical statistics and sampling theory to socio-economic statistics in the USSR at the end of the 1920s. It also reflects the inability of TsSU fully to overcome the methodological legacy of the Stalin period in the succeeding decades, which was discussed in Chapter A3.

Notes to Chapter B6

1 We have found no reference to representativeness checks for employees.

2 Postnikov (1953) describes a more elaborate procedure, with seven columns of average wages instead of four. Thus he includes a column of average wages of "all workers in the families selected who work in the enterprise selected". The advantages of including these additional columns, not mentioned in more recent sources, are unclear.

3 The exact value of the maximum permissible deviation is a matter of some confusion. Most frequently cited, for example by Matyukha (1967) and in United Nations (1978), is the range ± 3 - 5 per cent, with no explanation of how a range can be a maximum. Ananyeva (1964) gives ± 3 per cent, Venetskii and Matyukha (1968) ± 2 - 3 per cent, and Venetskii and Chernysheva (1978) ± 5 per cent. Grankov (1955) even mentions a "tolerance of 0.1", that is ± 10 per cent, for checking samples of collective farms. The value used may not be constant over time or even from place to place. According to Kozlov et al (1965 pp.230-1), the official instructions issued by TsSU USSR give no criterion of adequate representativeness except that "the indicators compared must be as close as possible". For convenience we shall assume that the limit is ± 5 per cent.

4 As a substitute for a representativeness check on per-capita income, Karapetyan, Rimashevskaya and Sidlyarenko (1967) propose that data on family size be collected for all workers on selection lists (from which workers who are not the basic worker of their family should be excluded), and that a check then be carried out on average family size as well as on average wage. They do not indicate the source of population data on family size. Separate checks on wage and on family size, however, do not amount to a check on wage per person if, as seems likely, the two variables are correlated.

5 Matyukha (1967) states that a check is made on the total wage of all family members working in industry, but it is hard to see how corresponding population data by branch could be obtained. Our interpretation of what is really meant as "the total wage of all family members working in the same branch as the selected worker" is in accordance with Postnikov (1953).

6 It is often recommended that "reasons" for significant discrepancies between sample and population averages be sought in failures to observe the sampling procedure laid down (Ananyeva 1964, Matyukha 1967). The search must often be fruitless, for observation of the sampling procedure by no means guarantees the absence of significant discrepancies.

7 According to Sidlyarenko, Head of the Budget Survey Department of TsSU RSFSR, speaking in 1964, the annual representativeness checks apply only to the collective farmer survey, while checks of the worker survey are carried out only about once every three years (Soveshchanie.... 1964).

8 Matyukha (1967) does describe, in a rather confused fashion, representativeness checks based on calculation of sampling error, but we conclude, in the light of all other evidence, that this is not intended as a description of actual practice.

9 As opposed to the official "social groups" (see Chapter B2).

10 For further examples of such checks, as applied to the results of a sample survey of outlays in retail trade in food products in 1964 and to those of a sample survey of railway freight operation sheets in 1961-2, see Volodarskii (1969).

CHAPTER B7

GENERAL ASSESSMENT OF THE REPRESENTATIVENESS OF THE BUDGET SURVEY SAMPLE

1 Introduction

In this chapter we attempt, to the extent that this is possible with the information available, a general assessment of the representativeness of the budget survey sample and of the data collected from it. We start with a summary review of the sources of bias discussed in Chapters B1 - B5 (Section 2). We then consider the net effect of all these biases on the single most crucial characteristic of the sample, its income distribution (Section 3). An assessment of survey expenditure data, by comparison with data from Soviet statistics of retail trade, follows (Section 4). This puts us in a position to draw conclusions (Section 5).

2 Review of sources of bias

We bring together all the sources of bias in the sample of any importance of which we are aware in Table B7.1, with indication of their effects on the sample income distribution and on other sample characteristics where known. Almost all biases are associated either with the basic methods of selecting families for the sample (such as the "tail-cutting" and "multi-worker" biases), or with incomplete coverage of the working population by territory, branch, skill category etc., or with the extended periods of participation of families in the survey.

If we take into account those biases in survey data resulting not from biases in the sample but from inadequate primary records - negligent record-keeping by families, concealment of expenditure on alcoholic drink from the interviewer etc. - then we see that the data of the budget survey are subject to about twenty different biases, most of them major in nature.

TABLE B7.1 - SOURCES OF BIAS IN THE BUDGET SURVEY SAMPLE(1)

No.	Nature of bias	Ref. Chapter/ Section	Effect of bias on: Sample income distribution	Effect of bias on: Other sample characteristics
1	"Tail-cutting bias" due to systematic sampling with mid-interval starting-point	B1/5	Reduces dispersion	
2	"Multi-worker bias" due to non-correspondence of selection unit (individual worker) with unit of observation (family): probability of inclusion depends on number of working family members	B1/6	Increases mean	Increases mean family size and mean worker-dependent ratio (5)
3	Non-coverage of households without working members (eg. student households, disabled households) except for recent coverage of some types of old-age pensioner household	B4/2	Increases mean	Increases mean family size
4	Non-coverage of branches of the economy employing about 40 per cent of all workers and employees - eg. most food industry, forestry, non-rail transport, communications, trade, public catering, everyday and residential services, non-school education, science, arts, finance, State and economic administration, material-technical supply, local industry	B4/2	Increases mean and reduces dispersion (2)	
5	Over-representation of heavy industry and under-representation of light industry and services	B4/2	Increases mean	
6	Non-coverage of enterprises with very small workforces (fewer than 50 or 100 workers), of craft cooperatives and of subsidiary collective farm enterprises	B4/2	Increases mean	Reduces representation of workers with a rural lifestyle (3)

cont'd...

TABLE B7.1 (cont'd)

No.	Nature of bias	Ref. Chapter/ Section	Effect of bias on: Sample income distribution	Effect of bias on: Other sample characteristics
7	Over-representation of leading enterprises in the main industrial centres; poor coverage of workers and employees in small towns	B3/7	Increases mean	Reduces representation of workers with a rural lifestyle (3)
8	Non-coverage of unskilled workers (grades 1 and 2 on the 6-point scale), of workers not on "basic activity" (eg. repair staff, security staff), of apprentices, temporary staff etc.	B4/6	Increases mean	
9	Non-coverage of collective farmer households without able-bodied members	B4/7	Increases mean	Increases mean family size, reduces mean age
10	Non-coverage of collective farmer households which share a cow with another household	B4/7	Increases mean	Increases mean livestock ownership
11	Non-coverage of over one-third of all territorial units at oblast' level, including sparsely populated areas in Central Asia (eg. the mountainous parts of Tadzhikistan), the North and Siberia (eg. all of Eastern Siberia except Primorsk krai)	B3/4	Unclear	Reduces representation of inhabitants of isolated areas with extreme climatic conditions (4)
12	Uneven coverage of territorial units in which the survey is conducted (with reweighting only of collective farmer data)	B3/5	Unclear	
13	Failure fully to reflect changes in social, economic and territorial structure of population (eg. expansion of "non-productive" sphere, urbanisation, appearance of new industrial regions, replacement of collective by State farms) in the sample	B1/7, B4/4	Unclear	Reduces relative representation of urban population, workers on State farms etc.

cont'd...

TABLE B7.1 (cont'd)

No.	Nature of bias	Ref. Chapter/ Section	Effect of bias on: Sample income distribution	Effect of bias on: Other sample characteristics
14	Over-representation of long-established households headed by older people and under-representation of new "young" households due to slow turnover of sample	B1/7	Increases mean	Increases mean family size and mean worker-dependent ratio (5)
15	Workers in sample have longer average period of service at their present enterprises than workers in the population as a whole, and this is associated with higher average levels of skill, wages and bonuses.	B1/7	Increases mean (5)	
16	Income of families in sample is raised by payment for participation and by other material incentives intended "to preserve a stable sample"; free entertainment, holidays etc., may also bias structure of consumption.	B5/3	Increases mean	
17	The "guinea-pig effect": long-term participants in the survey are thought to develop "regularised" budgetary behaviour atypical of the population as a whole (eg. higher rate of saving, reduced expenditure on "frivolous" goods).	B5/5	-	
18	Bias arising from "directed selection" of replacements for those who leave the survey from among families with average basic characteristics	B8/3	Reduces dispersion	Reduces dispersion of basic characteristics

cont'd...

TABLE B7.1 (cont'd)

Notes:

(1) Some biases in the data, eg. that due to concealment of expenditure on alcoholic drink (B5/5), are unrelated to bias in the sample.

(2) Because both branches with the lowest wages (eg. trade, everyday and residential services) and branches with the highest salaries (eg. science, State administration) are excluded.

(3) That is, who have private plots, privately constructed houses etc., - very common in the smaller towns as well as rural areas proper. This affects purchases made, as does the narrower range of consumer goods available compared to the main centres.

(4) Affects structure of consumption (clothing etc.).

(5) Biases nos. 2, 14 and 15 have mutually reinforcing effects.

Moreover, in important ways these biases tend not to cancel one another out but to be cumulative in effect. For example, no fewer than twelve of them (nos. 2-10, 14-16) tend to produce an upward bias in the average income level of the sample, while none of them clearly tends to produce any bias in the opposite direction.

Nevertheless, the biases interact in complex ways, so that often it would be impossible even to guess with any degree of confidence at their net effect on particular sample characteristics. This is so not only for non-Soviet observers but also, we judge, for Soviet researchers, including those with full access to Soviet statistical sources, and explains Soviet reluctance to use budget data for practical purposes. Thus we list in Table B7.2 some known or strongly suspected specific biases in the consumption data, but it would be very difficult to estimate to what extent they are exacerbated or counterbalanced by other biases in the sample, such as that of the income distribution.

Except for statements by officials responsible for the survey to the effect that the sample is "quite representative" (Matyukha 1962), there is a general Soviet view that the sample is "insufficiently representative" or "extremely unrepresentative". (e.g. Lakhman and Frenkel 1967, Berzkaln 1968, Planovyi... 1981 p.178). Our review certainly confirms this assessment.

3 Effect of the biases on the income distribution of the sample

As income level is the single most powerful determinant of consumer behaviour, Shvyrkov (1965) considers the correspondence of sample and population distributions of per-capita income the most appropriate criterion of overall sample representativeness. The comparisons between sample and population mean income levels and income distributions shown in Tables B7.3 and B7.4 confirm our expectation from Table B7.1 of a substantial

TABLE B7.2

SOME SPECIFIC KNOWN BIASES IN BUDGET SURVEY CONSUMPTION DATA

- downward bias in "petty" expenditures (eg. on matches, ice-cream) as a consequence of negligent record-keeping by families (Chapter B5/5)

- downward bias in expenditure on alcoholic drink and possibly on other socially disapproved products (eg. tobacco, cosmetics) as a consequence of concealment of expenditure by family from interviewer (Chapter B5/5)

- downward bias in expenditure by workers on equipment and supplies for private plot husbandry and private building as a consequence of the under-representation of workers with a rural lifestyle (biases 6 and 7 in Table B7.1)

- downward bias in expenditure on clothing and other products needed under extreme climatic conditions (bias 11 in Table B7.1)

- downward bias in expenditure on "frivolous" goods as a consequence of the "guinea-pig effect" (bias 17 in Table B7.1)

- upward bias in the rate of savings, but downward bias in stocks of ready cash, as a consequence of the "guinea-pig effect"

upward bias in the income levels of the sample. Moreover, these comparisons considerably understate the bias, because they use the incomes survey of 1958 as their source of data on the income distribution of the population, but this survey itself had an upward income bias: it was organised on the branch principle and was therefore subject to the "multi-worker bias", and it did not cover workers and employees in agriculture, who constituted 12 per cent of all workers and employees in 1958 (Rimashevskaya 1965 pp.60-61)[1].

Apart from the upward bias in average income, we expect three of the biases listed in Table B7.1 (nos. 1, 4 and 18) to reduce the dispersion of the sample income distribution below that of the population distribution by producing under-representation of the very highest income groups in the population as well as under-representation of the lower income groups. The intervals used for the distributions shown in Table B7.4 are presumably too wide to reveal an effect which emerges only at very high incomes, but both Berzkaln (1968) and Levin (1974 p.196) state that families with low incomes and families with high incomes are under-represented in the sample, while families with intermediate incomes are over-represented.

Korovkin (1969) reports that the number of "middle-income" families in the sample is ten times greater than the number of families with relatively high or low incomes. In Republic and oblast' samples families belonging to the "extreme" income groups are counted in tens or even ones, so that data relating to them cannot be used. For an example of this we may take the analysis by Golub (1976) of the per-capita income distribution of the sample of families of workers and employees in Donetsk oblast' in 1968. The use of 9-12 income groups would have been most suitable for his purpose of constructing a "differentiated balance of the incomes and consumption of the population" (see Chapter D2), but this gave only 20, 12 and 5 families respectively in three of the groups, so that he was forced to use just three income groups.

TABLE B7.3

BIAS OF FBS DATA ON AVERAGE FAMILY INCOMES OF WORKERS AND EMPLOYEES

Source:	Shvyrkov[1] (1965)	Shvyrkov and Aidina[2] (1968)
Indicator	Bias %	Bias %
Average family income from wages	+ 16	+ 18
Average family income from other sources	+ 32	
Average total family income	+ 16	+ 17

Notes: (1) Biases estimated by Shvyrkov (1965 p 27) by comparing FBS data with incomes data from the microcensus of 1958

(2) Biases estimated by Shvyrkov and Aidina (1968 p 245) by comparing uncorrected FBS data relating to workers' families in the RSFSR in 1963 with corresponding "corrected" data, obtained by reweighting the data with a more accurate income distribution (presumably taken from the microcensus)

TABLE B7.4

BIAS OF FBS DATA ON INCOME DISTRIBUTION

(a) Bias of income distribution of persons (family members) according to Shvyrkov (1965 p 27)

Groups in ascending order of per-capita family income	Ratio of proportion of persons in group according to FBS data to proportion in group according to data from 1958 microcensus
	%
1	32
2	71
3	106
4	147
5	177
6	187

TABLE B7.4

BIAS OF FBS DATA ON INCOME DISTRIBUTION

(b) Bias of income distribution of families estimated from the diagram given by Rimashevskaya (1965 p 60) (1)

Groups in ascending order of per-capita family income	Income distribution of families according to:		Ratio of income distributions according to FBS and microcensus data
	FBS data	Data from 1958 microcensus	
	(col. 1)	(col. 2)	(col. 3) = (col. 1)/(col. 2)
	%	%	%
1	4	17	23
2	16	25	63
3	35	33	107
4	28	17	158
5	17	7	238
	100	100	

Note:

(1) This diagram shows the income distributions according to the FBS and the microcensus. Its axes are not calibrated, and the Table is based on the measurements made of the diagram by the writer.

The population of families of workers and employees to which the diagram refers is not specified in the source. In conversation with the writer Rimashevskaya stated that, so far as she could recall, the diagram related to Armenia (Shemfield 1982b).

In fact, even if all income groups in the population were represented in the sample with exact proportionality, the very highest income groups constituting very small percentages of the population would be represented at oblast' level by a mere handful of families (assuming no change in the total sample size). Leifman (1963) therefore advocates that data be collected on a special supplementary subsample of families with very high incomes, so that the sampling fraction for this stratum would be considerably greater than that for the sample as a whole[2]. His concern that there be a fairly even distribution of the sample over a wide range of incomes is explained by his desire to derive estimated regression coefficients of consumption on income more precise than is possible using FBS data clumped together in only 6-10 per-capita income groups concentrated in the central part of the range (see Chapter D3).

Many of the Soviet economists engaged in the study of consumer demand are particularly interested in obtaining adequate budget data, and thus sufficiently precise estimates of regression coefficients, for the top income groups because they regard the structure of consumption of families in these groups as an approximation to the "rational" structure which it is assumed the rest of the population will adopt in due course. Information on top-income families is therefore needed for use in long-term demand forecasting (Korovkin 1969, Milner and Gilinskaya 1981). Thus Krylov (1970) determines "rational norms" by studying the consumption of families of Moscow workers and employees with per-capita incomes over 100 rubles per month.

One obstacle to the collection of more data about the very highest income group in the USSR is the fact that reference even to the existence of such a group is a politically sensitive matter. A rigorous censorship surrounds the subject, apparently determining the lower bound of the top open interval in any published income scale (Matthews 1978). It is

interesting that requirements for data arising out of one ideological doctrine - the assumption of a high future standard of living for the whole population - should come into conflict with another ideological myth - the absence of extreme contrasts in living standards in present-day Soviet society. The neglect of the lowest income groups in the budget survey may also, of course, be conditioned by ideological considerations.

4 Comparisons with trade statistics

Soviet economists, in discussing the representativeness of budget survey data on consumption, frequently refer to comparisons between these data and retail trade statistics based on the statistical reporting of trade establishments (State and cooperative trade). Some of these comparisons are set out in Table B7.5. They may relate to a limited territory, such as the figures for Perm oblast' presented in Table B7.7, but often they appear to relate to the USSR as a whole (as in Table B7.6). As these representativeness checks are not mentioned in accounts of the survey by TsSU officials such as Matyukha, we assume that they are conducted by economists in non-TsSU institutes on the basis of data obtained from TsSU[3]. We also assume that the various adjustments needed to make trade data as comparable as possible with budget survey data (see Chapter D3) are in fact carried out[4].

In Table B7.8 we show some comparisons by Shvyrkov and Aidina (1968) between budget survey consumption data and the same data when "corrected" by reweighting with the income distribution of the appropriate population, presumably taken from the 1958 incomes survey. The low levels of bias indicated by this method, by comparison with those indicated by the use of trade statistics, are consistent with the view that the reweighting under-corrects the data because of the upwards income bias in the income survey itself.

Zhutovskaya (1966) mentions an exercise carried out by TsSU on the

TABLE B7.5

BIAS OF BUDGET SURVEY DATA ON EXPENDITURE ON DIFFERENT GOODS AND SERVICES

Source	Method of estimation of bias	Bias	Items with greatest bias
Levkova and Tikhova (1968)	Comparisons with trade statistics of rural food sales in 1959-63; see Table B7.6	Up to 55%	meat, sausage, bread, groats, potatoes
Neshitoi and Shatalova (1980)	Comparison with trade statistics for Perm oblast' in 1967-8; see Table B7.7	Up to 80%	clothing, footwear, alcoholic drink
Shvyrkov and Aidina (1968)	Comparison with budget data for workers' families in RSFSR in 1963 reweighted by income distribution; see Table B7.8	Up to 25%	entertainment, clothing, meat, milk, eggs
Korovkin (1969 pp.113-4)	Comparison with trade statistics (further details not given)	Up to 60%	alcoholic drink
Bredov and Levin (1971)	"	Up to 30-40%	meat, sausage, potatoes
Levin (1973)	"	Up to 30%	(not indicated)

TABLE B7.6

DISCREPANCIES BETWEEN THE DATA OF RETAIL TRADE STATISTICS AND FBS DATA ON THE PURCHASE OF FOOD PRODUCTS IN STATE AND COOPERATIVE TRADE BY COLLECTIVE FARMER FAMILIES (1) IN 1959-1963

Product	Discrepancy (2) %
Bread	55
Meat	51
Potatoes	47
Sausage products	37
Groats	35
Flour	21
Vegetables	18
Macaroni products	17
Animal fats	12
Fish	8
Cheese	8
Herrings	5
Vegetable oils	3
Sugar	2

Source: Levkova and Tikhova (1968)

Notes: (1) As purchases made by collective farmer families in particular cannot be extracted from trade data, we assume that purchases in cooperative trade were used by Levkova and Tikhova as a proxy.

(2) The figures are given in the source as deviations of trade data from FBS data, with no indication of the direction of the deviations.

TABLE B7.7

COMPARISON OF THE STRUCTURE OF PURCHASES IN STATE AND COOPERATIVE TRADE BY THE POPULATION OF PERM OBLAST' (RSFSR) IN 1967-8 ACCORDING TO FBS DATA WITH THAT ACCORDING TO TRADE DATA (1)

Products and product groups	Proportion of ... of expenditure according to: FBS data (col. 1) %	Trade data (col. 2) %	Discrepancy of FBS data (col. 1) compared with trade data (col. 2) (col. 3) %
Bread and bread products	10·1	8·5	+ 19
Sugar and confectionery products	9·7	8·4	+ 15
Fish and fish products	2·6	2·3	+ 13
Meat and meat products	7·9	8·9	- 11
Fruit	2·1	2·4	- 12
Animal fats	1·6	1·9	- 13
Milk and milk products	3·3	4·0	- 18
Eggs	0·8	1·0	- 22
All food products except alcoholic drink	38·2	37·4	+ 2
Alcoholic drink	12·0	18·5	- 35
Tobacco products	1·4	1·7	- 20
Alcohol and tobacco	13·4	20·2	- 34
Clothing	14·6	8·2	+ 78
Footwear	6·1	4·5	+ 37
Knitwear	4·1	3·3	+ 25
Headwear	0·22	0·18	+ 22
Cloth	3·2	3·0	+ 7
Socks and stockings	1·1	1·1	0
All clothes and cloth	29·3	20·2	+ 45
Electrical products	1·30	1·02	+ 27
Watches and clocks	0·33	0·29	+ 14
Furniture	2·40	2·33	+ 3

TABLE B7.7 (cont'd)

Products and product groups	Proportion of expenditure according to: FBS data (col. 1) %	trade data (col. 2) %	Discrepancy of FBS data (col. 1) compared with trade data (col. 2) (col. 3) %
Radio products	1·55	1·62	- 4
Miscellaneous products (2)	3·10	3·25	- 5
Musical products	0·10	0·16	- 38
All industrial products (3)	8·8	8·7	+ 1
Other products (4)	10·3	13·5	- 24

Source: Neshitoi and Shatalova (1980), Table 5

Notes: (1) The product nomenclatures of the two data sources were adjusted to make them comparable.

(2) Items of household use, cultural products not included elsewhere, notebooks and stationery, sports products and toys

(3) All non-food products except tobacco products, clothing and cloth

(4) Products for which comparable categories could not be extracted from the nomenclatures of the two data sources - for example, certain food products, cars and bicycles, consumer durables and haberdashery.

TABLE B7.8

BIAS OF FBS DATA ON AVERAGE FAMILY EXPENDITURES (1)

Item of family expenditure	Bias %
All food products	+ 13
All clothing	+ 17
Theatre, cinema and other entertainments	+ 23
Bread and bread products	- 1
Potatoes	+ 1
Fish and fish products	+ 12
Meat and lard	+ 16
Milk and milk products	+ 16
Eggs	+ 17

Source: Shvyrkov and Aidina (1968 p 245)

Note: (1) Data relate to workers' families in the RSFSR in 1963. The biases were estimated by comparing FBS indicators with corresponding "corrected" FBS indicators, obtained by reweighting the data with a more accurate income distribution (presumably taken from the microcensus).

basis of income distribution data from the incomes survey of 1965 to reweight budget survey consumption data in a similar way. It was found that the reweighting had a much smaller effect on data for basic foods (eg. bread, potatoes, cereals, milk, sugar) and for such necessities as cheap fabrics than it did on data for more expensive foods (eg. butter, cheese, smetana, eggs, fruit) and non-food goods (eg. woollen and silk fabrics)[5]. This distinction is consistent with the figures derived in the similar exercise of Shvyrkov and Aidina (1968), as well as with the results one might expect from an upward income bias.

The distinction does not, however, seem consistent with our information about the comparisons with trade statistics. In the data of Levkova and Tikhova (Table B7.6), bread and potatoes - basic foods - appear together with meat and sausage products as the items with the highest bias. As the direction of bias is not indicated, perhaps the bias for bread and potatoes is a downward one. Bredov and Levin (1971) likewise refer to meat, sausage products and potatoes together as the items with the greatest bias. In the data for Perm <u>oblast'</u> (Table B7.7), the pattern of biases for basic and non-basic foods certainly does not conform to the expected impact of an upward income bias, while the pattern for non-food goods (eg. clothing and footwear) <u>is</u> roughly consistent with expectations. It might be that special characteristics of the budget survey sample at <u>oblast'</u> level can cause an unpredictable and otherwise inexplicable pattern of bias. This could also account for the higher overall level of bias in the Perm sample as compared with that at USSR or RSFSR level[6].

5 Conclusions

The FBS sample is extremely unrepresentative of the Soviet population as a whole. In particular, it has a severe upward bias in its average per-capita income level, although families with very high incomes as well as

families with low incomes are under-represented. However, one cannot even roughly predict on the basis of the income bias the overall pattern of biases in the data, especially at oblast' level. This is not surprising in view of the large number of different sources of bias at work in the sample, which must interact in complex ways. This unpredictability makes reliance on FBS data all the more problematical.

Notes to Chapter B7

1 Note that Tables B7.4(a) and B7.4(b) are not comparable because (a) is a distribution of persons while (b) is a distribution of families. So far as Rimashevskaya remembers in 1982, the diagram underlying Table B7.4(b) was based on data for Armenia only (Shenfield 1982b). If this is so, it (and thus the Table) are extremely imprecise, as the budget survey sample for Armenia was only about 150 families (Rimashevskaya 1965).

2 Leifman presumably envisages the necessary compensatory reweighting of data.

3 It is not clear just where the work is done. It is conceivable that it is done at economic research institutes, which are known to carry out independent data-processing on a large scale.

4 It is not possible to make retail trade data fully comparable with FBS data. For example, trade data does not cover private transactions.

5 Zhutovskaya reports quite a large discrepancy for expenditures connected with private plot husbandry and private building, explained in terms of the under-representation in the sample of workers with a rural lifestyle (bias No.7 in Table B7.1).

6 Some of the maximum biases listed in Table B7.5 are too low because expenditure on alcoholic drink, subject to very great bias, has been left out of consideration.

We have estimated the average bias corresponding to the figures in Tables B7.6 and B7.7, using the retail trade turnovers of the different products in 1963 as weights; the results were 44 per cent and 26 per cent respectively.

CHAPTER B8

EFFORTS TO IMPROVE SAMPLE REPRESENTATIVENESS

1 Introduction

Although the FBS sample has always suffered from severe biases, and although TsSU has always resisted carrying out the fundamental reconstruction of the survey necessary to eliminate these biases, certain efforts are made by TsSU to improve the representativeness of the sample. The top TsSU officials from time to time exert pressure on the TsSU Budget Statistics Department to do more in this regard:

> The Collegium (of TsSU USSR) noted that... measures had been taken to improve the representativeness and reliability of FBS data... The Department (of Budget Statistics) was instructed to ensure the reliability and representativeness of FBS data, and to render more active help in this matter to Republican TsSU's and local statistical bodies... (V kollegii... 1984b).

This chapter is devoted to considering the nature of these efforts, which are of two kinds. First, relatively large-scale reorganisations of the sample are carried out from time to time (Section 2). Second, an attempt is often made to improve the representativeness of the sample, as conceived of by TsSU staff, in the course of selecting replacements for those who leave the survey (Section 3).

2 Reorganisation of the sample

Known details of reorganisations of the sample which have been carried out since 1952 are shown, together with the dates of associated influential events, in Table B8.1. Our information is too scanty to make us confident that this is a complete list. For example, Korovkin (1969) refers to corrections to the sample "in recent years" - presumably 1965-8 - "in

TABLE B8.1

CHRONOLOGY OF KNOWN IMPROVEMENTS TO THE FBS SAMPLE

Events possibly or certainly influencing TsSU decisions	Improvements to the FBS sample
Mid to late 1950s. FBS data begin to be widely used for planning and administration, drawing decision-makers' attention to their deficiencies	
	1960: Sample expanded. Workers in construction, on the railways and on State farms included for the first time (ref.1).
	Early 1960s: Samples of workers in the building materials industry, in non-ferrous metallurgy and on State farms increased; samples of workers in the textiles and coal industries reduced (ref.2)
1966: Government discusses question of improving the FBS (ref.3)	
	1968-69: Workers in the food industry (or parts of it) included for the first time (ref.5).
1969: FBS severely criticised at an important conference on the study of consumer demand (ref.4)	
	1969-70: Sample expanded. Samples of workers in construction, on the railways and on State farms and of teachers, doctors and nurses increased. Territorial distribution of the sample of collective farmers corrected (ref.6).
	1977: Old-age pensioners included on a substantial scale for the first time (ref.7).

References:
(1) Krasnoshchekov (1962), Korovkin (1969), Matyukha (1969);
(2) Soveshchanie... (1964) - information relates to RSFSR;
(3) Vsesoyuznoe... (1969 pp. 24-25);
(4) Organizatsiya... (1971 pp.210-215, 245-6);
(5) Soveshchanie... (1969);
(6) Soveshchanie... (1969), Vsesoyuznoe... (1969 pp.24-25);
(7) Soloviev and Druker (1981), Shenfield (1982b).

order to take account of changes in the structure of production and in family composition" in the population, but we know nothing of any changes in this period. However, we do know a certain amount about the changes accompanying the only two exercises in substantially changing the total size of the sample - the expansions in 1960 and in 1969-70.

We lack the data to assess how much was achieved by each of the various reorganisations, though we can certainly agree with Korovkin that they "have not basically solved the problem". What we shall attempt to consider is the question: under what circumstances is TsSU motivated to carry out a reorganisation of the sample? Our discussion must be somewhat speculative in view of the paucity of evidence in the area of the internal politics determining the behaviour of TsSU. Certain pointers do, however, suggest that, when TsSU overcomes its inertia and makes changes, it does so more in response to external pressure from political leaders than prompted by autonomous professional judgement.

The speech of V.N.Starovskii, then Head of TsSU, at the conference of statisticians in June 1957, explained the need to make budget survey data more representative by reference to such pressure (Vsesoyuznoe... 1959 pp.23-24). Budget data had been used in the past only for "general discussions and illustrations", and consequently nobody had been concerned with how representative they were. But now the data were being used for practical purposes, for example in determining agricultural procurements, and so the leaders were taking a sceptical interest in them:

> Not long ago officials of the Belorussian Government and Gosplan had an argument about some livestock data from the budget survey of collective farmers - the meat and milk output norms for private plots used by TsSU. Some of the officials reproached the statisticians for providing inaccurate data, while others held that the norms were correct...

> When, discussing the meat balance of a Republic, the statisticians declare: 'So much meat is produced in the social sector, so much in the private sector, and we need to procure so much', then naturally people begin to take an interest: 'What are these budgets? How are they collected? Are they reliable? Is the sampling conducted scientifically?' This, of course, is true not only of Belorussian officials...

Starovskii called upon those statisticians who previously had taken a "cool attitude" towards the budget survey to recognise its great importance and to regard it not as a purely departmental but as an "all-State" matter.

In the 1950s budget data began to be used, together with data from other sources, in the compilation of balances of money incomes and expenditures of the population (Chapter D2). The account of recommended methodology by Margolin (1957) makes it clear that the budget survey was regarded as an unreliable source, to be resorted to only when more reliable sources were not available, for example for data on the income of collective farmers from private sales. FBS data on growth in the money holdings of the population were to be corrected "on the basis of expert assessment".

We may reasonably suppose then that the improvements in the sample in 1960 and in the succeeding period, and also the effort devoted at this time to the construction of corrective coefficients (see Chapter B9), were an attempt to placate other State bodies, especially Gosplan, and the State and Party leadership, who were using budget survey data for the first time and becoming aware of their deficiencies[1].

The improvements to the sample in 1969-70 were likewise preceded by external political pressure. At the conference of statisticians in April 1968 Starovskii revealed that the Government had discussed the question of improving the budget survey in 1966 (Vsesoyuznoe... 1969 pp.24-25). TsSU had been set the task of achieving territorial representativeness, for which steps were being taken, and the task of extending coverage from industry to other branches of the national economy. As a result the coverage of construction, railway transport and State farms began in 1969,

while that of health and education was increased (Table B8.1).

It is also possible that improvement at this time was to some extent a response to the severe criticism of the budget survey by economists at an important conference on the study of consumer demand held in September 1969 (Organizatsiya... 1971, eg. pp.245-6). Matyukha replied to this criticism in a rather evasive way, but with the assurance that from 1969 the sample would be constructed in a more representative fashion, both territorially and in terms of income distribution (pp.245-6).

If improvements to the sample have been in response to external political pressure, the question remains as to why the pressure has not been more constant or become stronger in the course of time. In fact, the absence of any substantial reorganisation of the sample since 1970 (except for improved coverage of old-age pensioners) points if anything to a weakening of the pressure for change. It could be that there was greater determination from other State bodies to have the survey improved in the 1950s, when they were discovering the unreliability of the data for the first time, and that since then they have become accustomed to the situation and have adapted their procedures to make the best of it, by avoiding the use of budget data where possible and making rough adjustments to them where their use could not be avoided.

3 Directed selection of replacements

Two alternative methods are used to select replacements for those families which, for whatever reason, leave the budget survey. The standard method mentioned in official instructions consists in finding a replacement family similar in important characteristics - occupation, skill level and wage of the selected worker, total income and composition of the family - to the family replaced (see Chapter B1). As we noted, the effect of this method is to perpetuate the biases which develop in the sample as a

consequence of the long periods of participation. The alternative method, by contrast,known as "directed selection", has the aim of improving sample representativeness. According to Rimashevskaya, its large-scale use in recent years has served substantially to improve the representativeness of the sample as a whole (Shenfield 1982b). Gavrilov, Head of the Budget Survey Department of TsSU Lithuania, reports that changes in sample representativeness by branch and category of workers are monitored, and families replaced by directed selection when necessary (Sukhoruchkina 1970).

Directed selection is resorted to when representativeness checks show sample averages deviating by more than the tolerated percentage from the corresponding population averages (Chapter B6; Venetskii and Chernysheva 1978). Members of the sample the basic characteristics of which deviate the furthest from population averages may then be dropped from the survey and replaced by new families with characteristics close to the population averages. Families who leave the survey for other reasons are likewise replaced by such new families. As a result the deviations between sample and population averages are reduced until the representativeness checks are passed.

An elaboration of the method of directed selection is advocated by Kupriyanov, Head of the Budget Survey Department of TsSU Moldavia (Sukhoruchkina 1970). He argues that it is more effective to select replacements whose wages are close to likely <u>future</u> average wage levels, as given by plan data or by extrapolation of past growth in average wages.

Directed selection improves sample representativeness in the very restricted sense in which this is understood by TsSU officials - that is, it reduces the bias in sample means of the "basic" indicators for the subpopulations covered by the survey. But at the same time it does further harm to other aspects of sample representativeness. The distributions of the basic indicators are distorted by the systematic exclusion from the

sample of "extreme" units; variability plays no part in the concept of representativeness being used. As there is no control on the values of non-basic indicators for either replaced or replacement units, selected by the discretion of TsSU staff without any provision for randomisation, the consequences for representativeness on these indicators are quite unpredictable.

As Shvyrkov (1965) concludes, the method of directed selection

> ... not only does not help to raise the representativeness of the data as a whole but disorients the investigator. It is necessary to do away with such 'corrections'. They can greatly harm the management and planning of the national economy, while creating the appearance of high representativeness on selected budget indicators.

4 Conclusions

Under external political pressure TsSU occasionally expands and reorganises the FBS sample. This entails a certain measure of improvement, in particular the extension of coverage to various population groups previously excluded, but the changes are neither radical nor frequent enough to achieve an acceptable level of representativeness. Directed selection of replacements is also widely used in order to improve sample representativeness, but as the methodology is based on a restricted concept of "representativeness" its net effect is to substitute certain types of bias by others, in an uncontrolled and largely unpredictable way. The efforts made by TsSU to improve the sample cannot therefore be regarded as adequately successful.

Notes to Chapter B8

1 Users of FBS data come to suspect their reliability when they are aware of the deficient survey methodology, when they are familiar with other data sources giving discrepant results or when they have personal experience of the subject-matter which contradicts the data. Biases will then very often be widely known, though it may also happen that a bias in the data remains unknown for a long time. For an example of the latter concerning the position of the disabled, see Chapter D1.

CHAPTER B9

USE OF CORRECTIVE COEFFICIENTS TO ADJUST FBS DATA

1 Background and introduction

At the 1957 conference of statisticians there were calls, in particular from A.Kh.Karapetyan, to obtain information on the income distribution of the population by including a question on family income in the schedule of the 1959 population census. One of the main reasons advanced for doing this was that knowledge of the income distribution of the population would make it possible to assess and improve the representativeness of the FBS (Vsesoyuznoe... 1959 pp.244-7). P.G.Pod'yachikh, Head of the TsSU Population Census Administration, replied that the census would give unreliable data on incomes, but did not dispute the need to collect such data (p.273). TsSU's response to that need was to institute a new sample survey of family composition, incomes and living conditions, sometimes conveniently referred to as the microcensus. Microcensuses were first conducted, covering about 250,000 non-agricultural families, in 1958 and then in 1967. From the third survey in 1972, the microcensus has covered the agricultural as well as the non-agricultural population with a total sample of about 300,000 families, and has been carried out every three years.

Karapetyan had argued in 1957 that "we need to know the income distribution of the population in order to reconstruct the FBS on the principle of representativeness by income". When the new incomes survey was introduced, however, senior TsSU officials such as Postnikov (1961) and Matyukha (1962) described its purpose not as the provision of information for reconstructing the FBS sample, but as the derivation of weights for correcting data from the existing sample. By presenting this procedure as a convenient and effective alternative to the reconstruction of the sample,

they justified their continuing failure to carry out any reconstruction:

> The FBS sample can be corrected only by its overall reconstruction, but this is an undesirable operation because it is expensive and data quality is worse in the first year for newly surveyed families. But maybe such reconstruction is not necessary? Thus we shall argue... In order to obtain a correct description every year of changes in the population without changing the sample, we must reweight FBS indicators grouped, for example, by family income... (Postnikov 1961)
>
> It is extremely difficult to eliminate the discrepancies (between the FBS and more reliable sources of data on the population - SDS) by more rigorous construction of the budget network, and there is no special practical necessity for doing so. The shortcoming may be overcome more simply... (Reweighting FBS data with corrective coefficients) would raise their reliability manyfold, making it possible to use them in practical calculations (Matyukha 1962).

In this chapter an assessment is made of Matyukha's claim by examining the methodologies which have been developed for applying corrective coefficients to FBS data and the practical use which has been made of them. In Section 2 we explain the theoretical principles suggested for this work by Karapetyan and Rimashevskaya. In Section 3 we discuss the problems encountered when attempts are made to apply these principles in practice. In Section 4 we review the work which has actually been done in this area since 1958. We comment in the concluding Section 5 on the overall value of corrective coefficients as a method of improving FBS data.

The use of corrective coefficients based on trade statistics is not discussed in this chapter because it is dealt with elsewhere: see Chapter D3 on the use of such coefficients to correct demand forecasts based on FBS data, and Chapter B7 on Soviet assessments of FBS sample representativeness by comparison of FBS data with trade statistics.

2 The theory of corrective coefficients

A generalised procedure for extrapolating from an unrepresentative sample to the population was set out by Karapetyan (1964a,b) and further

explained by Rimashevskaya (1965). Omitting Karapetyan's algebraic analyses of alternative static and dynamic variants of his model, we here explain just the basic principles involved. There are two of these: that of modelling the variables under study as functions of population structure (modeli v funktsii struktury), and that of determining the population structure with the aid of a series of embedded data sources analogous to multiphase sampling.

The aim in "modelling as a function of structure" is to stratify both the sample under correction and the population into corresponding sets of groups, each of which can be assumed sufficiently homogeneous with respect to the indicators being studied. Those factors thought to exert the strongest influence on the indicators under study are selected as stratification variables. For each homogeneous group, the average sample values of the indicators under study are calculated. These average values ideally constitute a system of stable coefficients or "factorial norms" which can be applied to the structure of the population (that is, its composition in terms of the homogeneous groups) to give reliably extrapolated values of the indicators under study.

For example, Karapetyan considers that the following four factors are sufficiently important as influences on consumer demand to be used as stratification variables:

A. Natural-geographic zone (northern, middle, southern);
B. Type of settlement (urban, rural);
C. Social group (workers and employees, collective farmers); and
D. Level of material provision of family (high, middle, low).

These factors define 36 (that is, 3 x 2 x 2 x 3) homogeneous groups in the sample, for each of which a set of factorial norms of average consumption would be calculated for extrapolation from the structure of the population.

It is assumed in this structural modelling that the structure of the

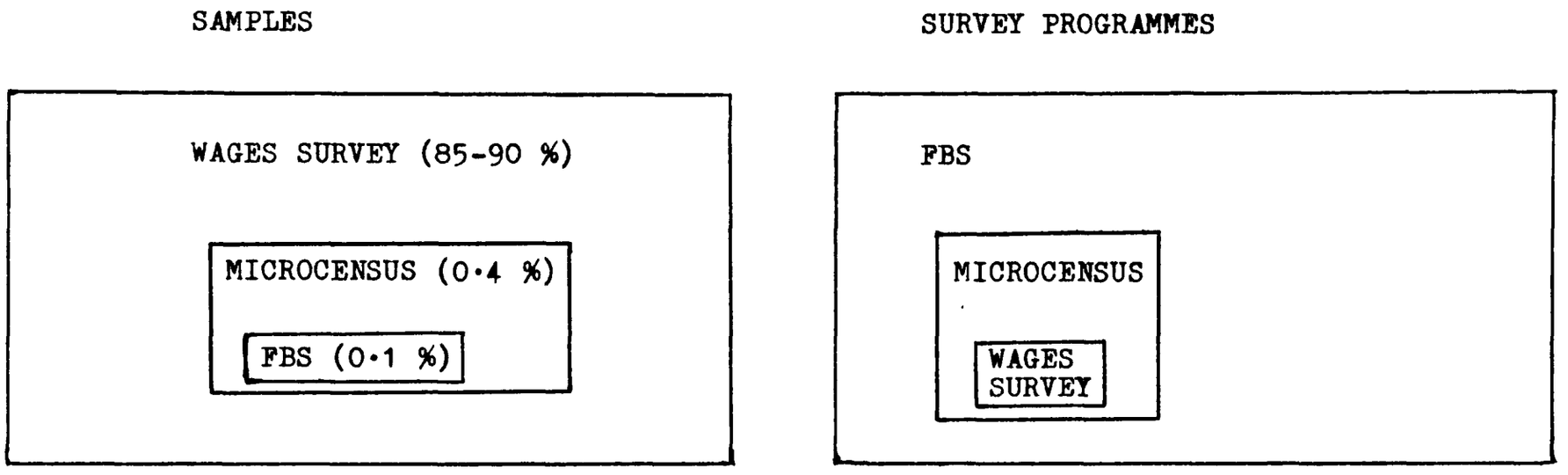
population is already known. The procedure which takes account of the prior need to assemble structural information in the required form is referred to as the "method of composition of data" (metod kompozitsii dannykh) from different sources.

It is argued that as a general rule a series of related data sources, embedded within one another will be available. There will be some sources based on observation of most or all of the population (censuses, statistical reporting etc), while other sources will be based on observation of larger or smaller samples. The smaller the number of units observed, the more extensive the survey programme but the less reliable the data obtained. The weaknesses of one type of data source are then fortunately offset by compensating strengths in another type of source. More reliable data on a narrow programme of observation can be used to correct less reliable data on a broad programme.

As an illustration we may take the series of three data sources used by Rimashevskaya to correct FBS data for workers and employees (Table B9.1). The microcensus was introduced to collect information on a range of questions more restricted than that covered by the FBS, but from a larger and more representative sample. In turn, the wages surveys collect information on wages alone but cover almost all of the the employed population.

Let us assume a series of k data sources: sources 1,2, ... (k-1) provide structural information on the population, while source k is the unrepresentative sample survey the data from which are to be corrected. Source 1, which comes closest to covering the whole population, provides initial structural information about the population on a very narrow set of indicators (perhaps on just one indicator). Source 2 covers a smaller sample, for which it provides information on the same set of indicators as source 1 and also on some additional indicators. By comparing the distributions of the shared indicators given by sources 1 and 2, it is

MULTIPHASE SAMPLING SYSTEM USED BY N M RIMASHEVSKAYA TO CORRECT FBS DATA (WORKERS AND EMPLOYEES)



FBS : Family Budget Survey, continuous, about 0·1 per cent of families

Microcensus : Survey of family composition, incomes and living conditions, about every three years, about 0·4 per cent of families

Wages survey : Survey of wages, every two years, about 85-90 per cent of employed workers and employees

Source: Rimashevskaya (1965)

possible to use cross-tabulations between shared and additional indicators from source 2 to "correct" the additional indicators from source 2. Source 3 is similarly used to "correct" source 2, and so on. When source (k-1) is "corrected", the structural information on the population is as reliable as the set of available sources permits. The "corrected" source (k-1) can now be used to "correct" the indicators under study as given by source k.

Rimashevskaya identifies the method of composition of data with multiphase sampling, a survey procedure in which successive subsamples are observed on successively broader programmes. For example, a census may collect some basic data on the whole population, some additional data on a 20 per cent sample, and still broader ranges of information on a 10 per cent sample contained within the first sample and then on a 5 per cent sample contained within both of the first two samples. Lenin had a high opinion, we are told, of the multiphase sampling used by the statisticians of Penza guberniya before the revolution.

Although there are obvious parallels between multiphase sampling and the method of composition of data,it is scarcely appropriate to identify the two. The errors in data from a properly designed multiphase sample are only sampling errors, while the main errors in the surveys linked together by the method of composition arise from their faulty sample design. The greater unreliability of data from smaller samples is not an inherent principle of sample surveys but an accidental by-product of the inadequacy of Soviet practice. The whole theory of corrective coefficients is an expedient designed to cope with the poor methodology used by TsSU; had TsSU survey design been sound, there would have been no need to develop it.

3 The practice of corrective coefficients

Attempts to apply the theory of corrective coefficients to the correction of FBS data have to reckon with a number of practical problems. Here we consider the least tractable of these problems, those which arise from the shortcomings of the data sources available, even the "most reliable" of them.

For consumption to be modelled as a function of structure, it must be possible to stratify both the FBS sample and the population according to the required schema. The size of each homogeneous group within the sample must be large enough to yield sufficiently accurate factorial norms. However, Karapetyan (1964b) states that, because the FBS covers such a restricted section of the population, cross-tabulations of the required type cannot be obtained from it for his four-factor model. The FBS, it appears, is not even representative enough to provide a basis for its own correction; a proper correction procedure will be feasible only after the FBS has been improved!

In practice only Karapetyan's factor D, level of material provision, as represented by family per-capita income, has been used to derive corrective coefficients. The need to correct for bias even in such an influential aspect of family composition as worker-dependent ratio has been neglected.

The effectiveness of the method of composition of data is clearly dependent on just how reliable the "most reliable" data source is as the source of the initial structural information. TsSU's two-yearly wages survey has two main shortcomings as an initial source of information on incomes: (a) its incomplete coverage; and (b) its use of the individual worker or employee, rather than the family, as unit of observation.

(a) The data on wage-distribution from the wages survey leave out about 10-15 per cent of employed persons. Rimashevskaya (1965) explains

that the distribution of workers and employees by wage must be made more precise using control figures from statistical reporting on the total number of workers and employees and on the total wages fund. The numbers in different wage bands are "corrected" by shifting the boundaries of the original intervals upwards so that the total wage corresponds to the known total wage fund. The implicit assumption that those omitted from the survey have the same wage distribution as those included is not, however, justified, and there is no reason to believe that the wage distribution constructed in this makeshift way is very accurate.

(b) The distribution required is not that of workers and employees by wage but that of families by per-capita family income. Rimashevskaya (1965) claims that the first of these distributions can be transformed into the second "using the rigid dependence between the two distributions" - a very dubious proposition in view of the complexities of family composition.

Aganbegyan and Rimashevskaya (1961) and Davidovich and Nazarov (1968) give a fuller account of how the wage distribution from the wages survey should be transformed into a family income distribution. The proportions of workers and employees at different wage levels who are the main earners ("first workers") in their families are known from the microcensus, and these proportions are used to derive from the overall wage distribution a wage distribution of first workers. Microcensus data on the numbers of families with one and with two workers, and on the wage distribution of second workers for each wage level of the first worker, are then used to obtain a distribution of families by the total wage of the first and second workers. Data on the ratios of total wage of first and second workers to other family income make it possible next to estimate a distribution of families by total family income. From tabulations of total family income against family size, one finally arrives at a distribution

of families by per-capita family income. This distribution is corrected on the basis of checks against statistical report data on the total wage fund, total income from other sources and so on.

This roundabout method cannot give very accurate results because of the large number of approximations which have to be used and, above all, because of its heavy reliance on data from the microcensus, not very much more reliable than data from the FBS itself (both being designed on the branch principle). Moreover, it takes no account of households without any wage-earners.

Alternatively distributions can be fitted to log-normal models,with the differential between the variance of logarithms of wages and the variance of logarithms of family incomes being estimated from microcensus data (Potrebnosti...1979 pp.186-7). The problem of reliance on the microcensus by a procedure designed to "correct" the microcensus remains.

A further problem with the use both of the wages survey and of the microcensus as data sources is that, unlike the FBS, they are conducted only at intervals of a few years. R.T.Barsukova writes of the use by NII TsSU of regression models to project estimated family income distributions from the time of the last microcensus to the precent (Planovyi... 1981 p.190). Practical exercises in the application of corrective coefficients have usually avoided this problem by restricting themselves to the use of FBS data for the same period as a microcensus.

In the light of all these difficulties, it is not surprising that the improvements to FBS data achieved by the application of corrective coefficients are generally described in moderate terms. Rimashevskaya (1968b) reports that experimental calculations on data for 1958 "substantially increased" the precision of data relating to the USSR as a whole, but that results for particular regions (for example, Republics) were not positive. Zhutovskaya (1966), describing similar work on 1958

data carried out in TsSU USSR, states merely that reweighting by the estimated distribution of families of industrial workers by per-capita income brought FBS data "significantly nearer" to the microcensus data on the shared indicators.

4 Work carried out on corrective coefficients

The exercises carried out since 1958 to derive corrective coefficients and apply them to FBS data, generally described as "experimental calculations", are listed in Table B9.2. There may have been other exercises not mentioned in the available literature. The exercises all deal with budget data for workers, or for workers and employees; we have found no evidence of attempts to correct data for collective farmers.

The earliest work on corrective coefficients was apparently carried out by Rimashevskaya and her colleagues at the Scientific Research Institute of Labour (NII-truda) from about 1960, on FBS data for 1958 and then for 1961. The method of composition of data was used, with income distributions being derived from the wages surveys (carried out by TsSU in 1956, 1959 and 1961) and the 1958 microcensus. A standard programme was compiled to do the calculations on the "Ural-2" computer in the Main Computer Centre of Gosplan USSR[1].

Zhutovskaya (1966) describes how the data for 1958 were also corrected, but on the basis of the 1958 microcensus alone, by TsSU USSR. Corrective coefficients were derived by comparing distributions by per-capita monthly income[2] from microcensus data and FBS data. Tabulations of money income by per-capita-income group were obtained from FBS data and reweighted, and then compared with corresponding tabulations from microcensus data as a check on the reliability of the corrective coefficients. The reweighting, as we have noted, brought the FBS data "significantly nearer" to the microcensus data. The coefficients were, finally, used to reweight FBS

TABLE B9.2

EXERCISES IN THE CALCULATION OF CORRECTIVE COEFFICIENTS REFERRED TO IN THE LITERATURE

Institution at which work done	Data sources used to correct FBS	Budgets used in calculations (1)	Dates to which FBS data used relate	References
NII-truda	Wages survey and microcensus		1958, 1961	Rimashevskaya (1965)
TsSU USSR	Microcensus	Industrial workers	1958	Zhutovskaya (1966)
Probably TsSU RSFSR (2)	Wages survey	Workers in RSFSR	1963	Bredov (1966), Shvyrkov and Aidina (1968)
Latvian Division of NII TsSU	Microcensus	Workers and employees in Latvia	1972 (?) (3)	Ananyeva and Ivanova (1972)
TsSU (?) (4)	Microcensus		October 1978	Kildishev et al (1980 Ch. 13)
NII TsSU (5)	Wages survey and microcensus			Planovyi... (1981 p 190)

Notes: (1) Where not indicated, presumably workers and employees in the USSR as a whole

(2) Working group consisting of Shvyrkov, Lipkin and Bredov based at Gosplan and at TsSU RSF

(3) Experimental calculations reported underway in 1972, presumably using data for that year

(4) Unclear; context suggests TsSU

(5) Reported that NII TsSU have worked out methodology; no reference to actual calculations

tabulations of expenditure and consumption. Zhutovskaya remarks that work on methods of extrapolating FBS data to the population, by deriving optimal and stable coefficients for different indicators, had only just begun.

The use of corrective coefficients was part of the work of a group of economists at Gosplan and TsSU RSFSR who, in the early 1960s, devised a programme of computer processing of family budgets by mathematical methods such as regression (Bredov 1966). Per-capita-income distributions of families were to be derived from the wages survey, and checked by comparing the total income implicit in a distribution with corresponding data from the national monetary balance. The distributions were to be used to correct average budget indicators for different regions. Shvyrkov and Aidina (1968) report a similar exercise, but one relying on an income distribution of family members rather than of families, carried out on workers' budget data for the RSFSR in 1963. Some of the results have already been presented in Table B7.1.

Plans for the computerisation of FBS data processing provided for the calculation and application of corrective coefficients (report by Bredov, Soveshchanie... 1964; speech by Matyukha, Soveshchanie... 1969). At a TsSU staff meeting in April 1970, Matyukha reported that the methodology of extrapolating FBS data to the population was being reworked in connection with the current expansion of the sample; the new methodology would be incorporated into the computer programme about to be introduced (Sukhoruchkina 1970).

However, the computerisation programme actually implemented at the beginning of the 1970s did not involve any use of corrective coefficients. Siskov, Klimenko and Ivanov (1978) and Ananyeva, Ivanova and Khlopina (1980) list the construction of income distributions of the population and their use in extrapolating FBS data to the population as tasks yet to be tackled as part of the second stage in the development of the Automated

System of State Statistics (see Chapter C2). The diagram of the planned computer subsystem for budget statistics, given by Ananyeva and Barsukova (1981), contains the two linked tasks "statistical study of incomes" and "extrapolation of budget data to the general population". It is not clear when these parts of the subsystem are to come into operation.

Ananyeva and Ivanova (1972) report, by way of exception to the situation at the level of the USSR as a whole, that the Latvian Division of NII TsSU do correct FBS data for Latvia, by fitting microcensus data to a log-normal model of income distribution.

The fact that work on corrective coefficients was in abeyance in the 1970s is reflected in the changing rationale attributed to the microcensus. While authors like Venetskii and Matyukha (1971) were still explaining the microcensus primarily as a means of extrapolating FBS data to the population, the account of the 1975 microcensus by Vorontsova (1975) does not even mention this as a purpose of the survey, the most important function of which she gives as "investigating the structure and formation of incomes".

The failure of the computerisation programme to date to apply corrective coefficients is obviously related to its failure to produce tabulations by per-capita income, to which the coefficients would be applied. These tabulations have not been produced since 1970 (Planovyi... 1981, p.183; Chapter C3). This would explain why the work on corrective coefficients in NII TsSU, as described by Barsukova (Planovyi... 1981 p.190), appears to be confined mainly to the development of methodology[3].

Kildishev et al (1980 Ch.13) nevertheless give an account of work to correct FBS data for October 1978 on the basis of data from the microcensus of September 1978. A crude method was, however, resorted to which avoids the use of income distributions. By comparing data from the two sources on shared indicators - family composition, money income and its composition - coefficients were obtained for recalculating other FBS indicators - money

expenditures and their composition, quantities consumed of food and non-food products and so on. At best, this can have amounted to pro-rata grossing-up of expenditure and consumption data from the FBS to correspond to average income data given by the microcensus for different family types. That is, the absence of income distribution data must have made it necessary to ignore the dependence of expenditure patterns on income level.

5 The value of corrective coefficients

As we have seen, the use of corrective coefficients based on other Soviet data sources can significantly improve the accuracy of FBS data[4]. It is even possible that a fairly effective correction procedure may be incorporated into the programme of computerised data-processing in the near future. There are, however, many problems which prevent such correction from yielding sufficiently reliable data. Some of these are inherent to the nature of the task of correction itself, such as the difficulty of taking account of a wide enough range of relevant biasing factors. Others are products of the deficiencies of Soviet data sources - both the FBS and the sources from which corrective coefficients can be derived.

There is a tendency for Soviet authors to call for improvements in data sources in order to make more effective correction procedures possible. For example, the Head of the Latvian Division of NII TsSU urges that microcensuses should be conducted more frequently and regularly and with broader coverage, and should moreover collect data on consumption, expenditure and household property as well as data on incomes, family composition and living conditions. This should at least be tried out experimentally in one or several Union Republics (Berzkaln 1968, 1972). In effect,he calls for the microcensus itself to be turned into (among other things) a recurrent budget survey, so that more reliable corrective

coefficients can be calculated directly from expenditure data instead of indirectly from income data. However, even if this were done, the biases in the design of the microcensus sample would still suffice to prevent the derivation of reasonably accurate corrective coefficients.

The simplest remedy would of course be to eliminate the whole artificial problem of corrective coefficients by reconstructing the FBS on a sound basis.

CHAPTER B10

THE TERRITORIAL AND THE BRANCH PRINCIPLE OF SAMPLE DESIGN

1 Introduction

The Soviet administrative system as a whole, and within it also the statistical system, operates through the interaction of two different principles of organisation - the territorial principle and the branch (or production) principle. We provide a context for the discussion of these principles in relation to the FBS by considering their role in Soviet administration in general in Section 2 and in Soviet statistics in particular in Section 3.

The issue most regularly brought up by Soviet critics of the FBS (for example, by Levin 1969, 1973; Lakhman and Frenkel 1971; Karapetyan 1980) is the need to reconstruct the sample, which they describe as organised on the branch principle, on a territorial basis. This would make the sample design broadly similar to that of budget surveys in Western countries (see, for example, Kemsley 1968, 1969, Department of Employment 1979 for the UK; Bigata 1973 for France) and in some East European countries (such as Hungary and Bulgaria).

In Section 4 we consider the territorial and branch principles in relation to the Soviet FBS. We argue that it is an oversimplification to state that the survey is organised wholly on the branch principle, and that it is better described as organised on a combined territorial-branch principle. The role played by the branch principle is nevertheless great enough to justify the criticisms of it advanced by Soviet economists and mathematical statisticians, and we review their arguments in Section 5. In Section 6 we review the arguments advanced in defence of the existing

organisation of the survey by TsSU officials. The proposals for reconstructing the survey on a fully territorial basis are examined in Section 7.

The examination of these proposals demonstrates that they have far-reaching implications for Soviet statistics. This leads on to a consideration in Section 8 of the factors which seem so far successfully to have impeded reform in spite of the extremely strong case for it. Some conclusions are suggested in Section 9.

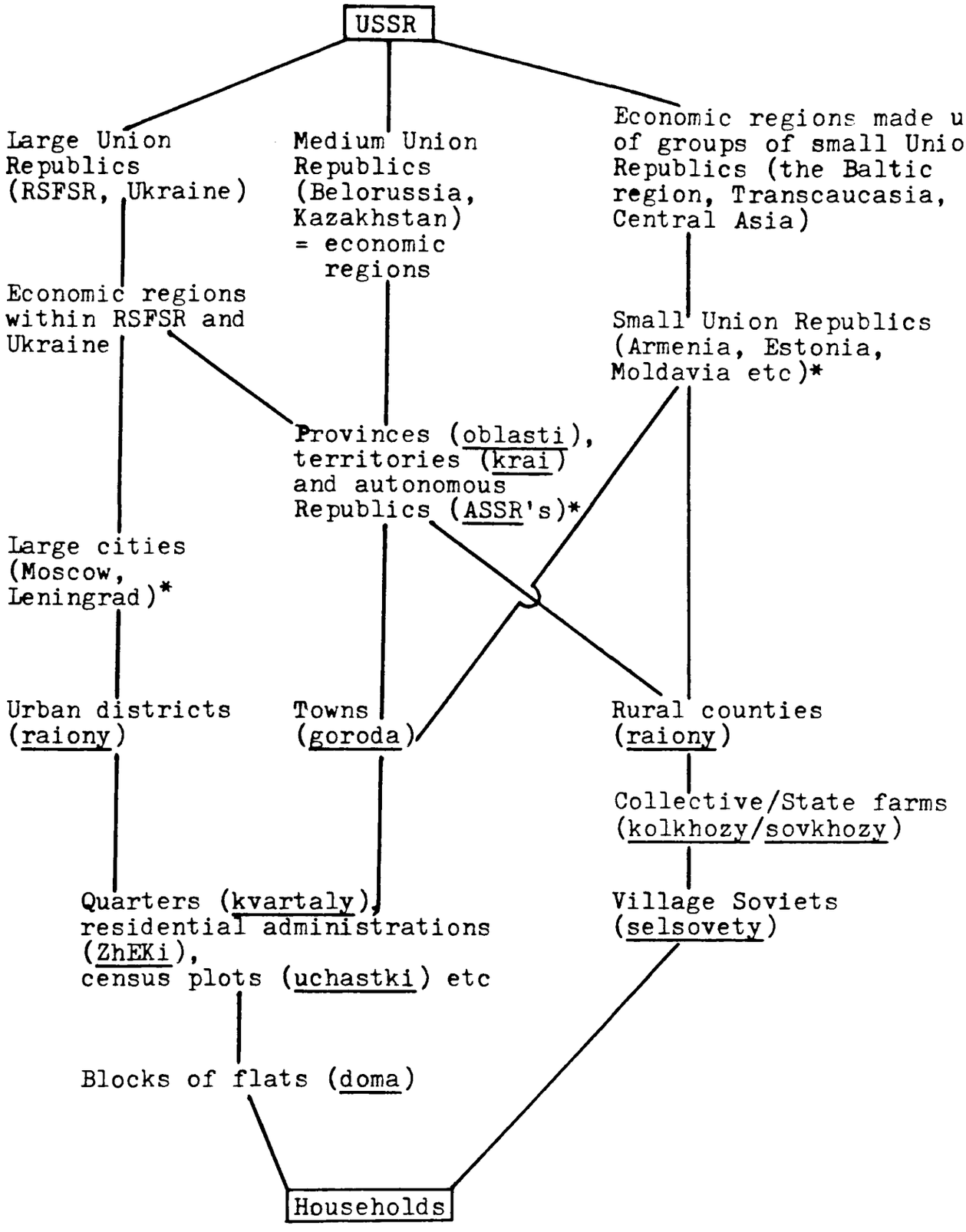
2 The territorial and the branch principles in the Soviet administrative system

The territorial principle in the Soviet system is represented by the hierarchies of Party committees and *apparats* and of Governmental ("Soviet") bodies at Republican, *oblast'*, *raion* and village levels responsible for general-purpose management of the corresponding territorial units. The rather complicated system of territorial units in use in the USSR is outlined in Table B10.1(a). The branch (or production) principle is represented by the hierarchies of central Ministries and establishments in Moscow, responsible for particular spheres of work activity throughout the territory of the USSR, and of their subordinate administrations. The branch principle is illustrated in Table B10.1(b).

When Soviet officials and scholars talk about their political system, reports Hough (1980 p.66), they frequently refer to the relations between "verticals" and "horizontals". The "verticals" are the Ministries and their lines of command, the "horizontals" the territorial bodies. Almost all local administrators are subject to "dual subordination" - that is, they are answerable both to vertical superordinates in their sphere of activity and to the horizontal authorities of the territorital units they work within. The relative influence over them of the two sets of

TABLE 10.1 - PRINCIPLES OF ORGANISATION IN SOVIET ADMINISTRATION

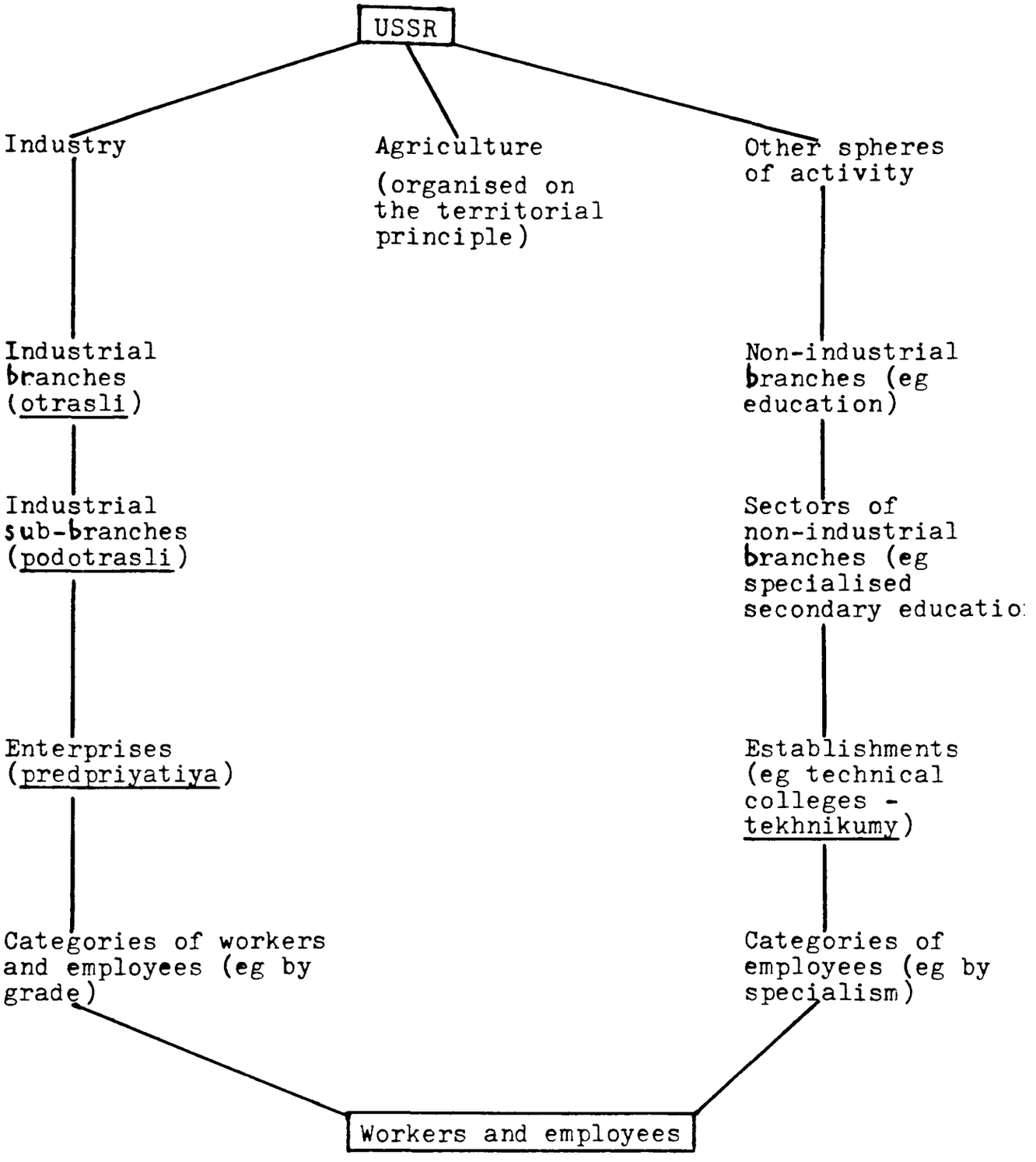
(a) THE TERRITORIAL PRINCIPLE (*territorialnyi printsip*)



* We refer to all territorial units immediately above raion level - large cities, oblasti, krai, ASSR's and small Union Republics (ie those Republics lacking subdivision into large cities, oblasti, krai and ASSR's) - as oblast'-level units.

TABLE B0.1 - PRINCIPLES OF ORGANISATION IN SOVIET ADMINISTRATION

(b) THE PRODUCTION OR BRANCH PRINCIPLE (proizvodstvennyi / otraslevoi printsip)



authorities varies over time and from one sphere of activity to another. In general, the strength of horizontal influences varies inversely with the priority assigned to the sphere of activity in the Soviet system. The horizontals make some effort to coordinate and influence for local purposes, such as town planning, the work of the many representatives of the verticals in their area.

However, it is in our view misleading to rely on a simple dichotomy between horizontals and verticals which ignores the great differences in the position of horizontals at higher and lower levels of the territorial hierarchy. The higher horizontals, at Republican and oblast' levels, have very much more power and play a much more central role in the Soviet political and economic system than the lower horizontals, at raion, town and village levels.

The conflict between verticals and higher horizontals is one between parties whose power is of a similar order of magnitude. Both Ministerial officials and Secretaries of oblast' Party committees (and of Republican Central Committees)[5] constitute substantial blocs in the Communist Party Central Committee, which according to Yanov (1977) has become the "Parliament" of the Soviet system in the post-Stalin period. Officials below oblast' level, on the other hand, enjoy no influence at the All-Union level. The contest between the lower horizontals and the verticals is a very unequal one.

Effective town planning, for example, is possible only in large cities like Moscow and Leningrad which have high-ranking Party organisations. Elsewhere the town authorities lack the power to stop the industrial Ministries pursuing industrial development with scant regard to the needs of town planning (Peers 1981). The Ministries assume functions which it would be more rational and equitable to assign to local authorities, such as the provision of housing and childcare facilities for their staffs.

The problem is especially severe in areas dominated by a single industry, like the oilfields of Western Siberia. There have been complaints in Pravda that the layout of the town of Tyumen' has become chaotic because the town Soviet is unable to control the building activities of the oil and gas organisations.

In the Soviet equivalent of a company town, the main local industrial enterprise may be the most important institution and its Director in practice the most powerful person, and one whose main responsibility is to his Ministerial verticals. Dunham (1976) cites the portrayal of this situation in a Soviet novel:

> For all practical purposes Rotov (Director of a steel combine) was also boss of the city. This was so because absolutely everything that was built in the city belonged to the plant. The plant ran the streetcar system. The Community Department of the plant had surfaced the streets. The Recreation Department of the plant had planted trees along the boulevards and in the parks. And the Chairman of the City Soviet did not usually come to Rotov to make demands as a superior. He came to make requests (Popov 1949).

The plant Director in another novel says of the First Secretary of the raion Party Committee:

> He is afraid to fight me... His raion is a poor one... The whole of its economic base is in my hands (Dudintsev 1957, quoted by Hough and Fainsod 1979 pp.506-7).

The local authorities, weak and without adequate resources of their own, are dependent on local industrial enterprises for all sorts of assistance - for the provision of equipment for a school, of spare parts for a collective farm, of temporary labour to help bring in the harvest or complete a construction project on time (Hough and Fainsod 1979)[1].

Ever since the end of the 1920s and with the exception of a few years under Khrushchev, economic planning and administration in the USSR have been conducted primarily (agriculture being a special case) through the

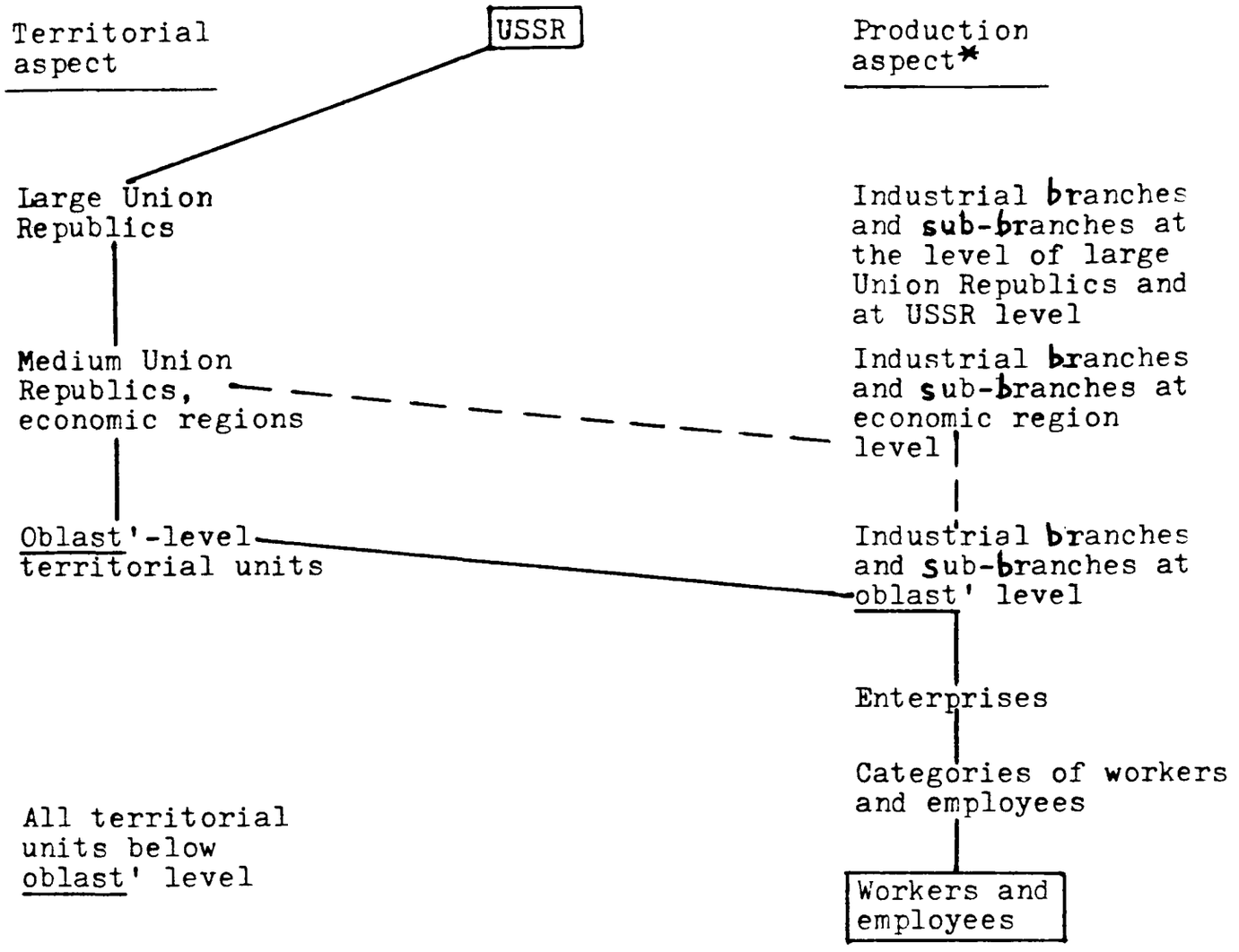
Ministries and the State Planning Commission (Gosplan) on the branch principle. Under Khrushchev most branch Ministries were abolished and their functions taken over by regional economic councils (sovnarkhozy), at first at oblast' level and later covering larger regions. The sovnarkhoz system did not, however, totally replace the branch by the territorial principle in economic planning, in that enterprises within the jurisdiction of any one sovnarkhoz were administered not by territorial unit (say, by raion) but by branch. That is, the territorial principle prevailed at higher levels - from oblast' level upwards, or from the level of economic region upwards - while the branch principle prevailed at lower levels. This combined principle of organisation, shown in Table B10(c) is known as the territorial-production or territorial-branch principle.

The abolition of the sovnarkhozy and the restoration of the Ministries brought the branch principle into a dominant position once more, but the territorial-branch principle has also retained considerable importance in the planning system[2]. A role is played in it by the oblast' planning commissions, and the territorial aspect of planning has been given increased attention in recent years with the development of integrated "territorial-production complexes". The oblast' Party authorities take a particularly active part in supervising plan fulfilment of enterprises in their areas, in facilitating supplies and the like (Yanov 1977). As we shall see, the statistical system has also continued to be organised on the territorial-branch principle.

The partial use of the teritorial-branch principle brings only the higher horizontals into the process of economic planning and administration which dominates Soviet society. The lower horizontals remain excluded, and this largely explains their lack of power and resources. The oblast' Party committee keeps in direct contact with the most important enterprises in the oblast', by-passing the raion committees. The functions of

TABLE 10.1 - PRINCIPLES OF ORGANISATION IN SOVIET ADMINISTRATION

(c) THE TERRITORIAL-PRODUCTION OR TERRITORIAL-BRANCH PRINCIPLE (territorialno-proizvodstvennyi / territorialno-otraslevoi printsip)



* The Table applies to industry. Similar patterns would apply to other spheres, except those that in their nature are organised on a territorial basis (agriculture etc).

The more usual system in which the switch from the territorial to the branch principle occurs at oblast' level is indicated by continuous lines, while another possible variant is indicated by broken lines.

territorial bodies below *oblast'* level are restricted to the provision of local services which tend to be denied high priority.

3 The territorial and the branch principles in the Soviet statistical system

Central Ministries and establishments collect their own "Departmental statistics" through their subordinate administrations on the branch principle. The Central Statistical Administration attempts to limit the collection of Departmental statistics, claiming a near monopoly position for the "State statistics" it itself collects.

Production statistics play the preponderant role in the work of TsSU. The only major exercises undertaken by TsSU which are unrelated to production are the population census and the budget survey. TsSU has resisted pressures to assume responsibility for other fields of statistics, preferring to leave health statistics in the hands of the Ministry of Health, statistics of pollution in the hands of the Sanitary-Epidemiological and Hydrometeorological Services (Motkin 1977), market research and public opinion surveys in the hands of special institutes and so on. The neglect of social statistics by TsSU often comes under criticism (eg. Rimashevskaya 1968b). Social scientists studying rural development comment on the imbalance in the types of data collected:

> Some aspects of village development (eg. the development of agricultural production or the employment structure of the population) are covered by hundreds and thousands of different statistical indicators, while other aspects (degree of social activism of the population, the state of its health,the development of cultural consumption etc.) are either reflected in a small number of indicators or not recorded in statistics at all (*Sotsialno-demograficheskoe* ... 1980).

The collection of production statistics by TsSU is organised on the territorial-branch principle. Almost all the statistical reports of enterprises are submitted, not through the local town and *raion*

Inspectorates of State Statistics, which are very small offices of peripheral importance, but directly to the appropriate Departments of the oblast' statistical administration, which process data on the branch principle. From the oblast' offices summary data are passed up the territorial hierarchy of TsSU offices to Republican and thence to All-Union level.

As production itself is administered on the branch and territorial-branch principles, it is to be expected that production statistics should also be so organised. Demographic and social statistics, however, are almost exclusively concerned with populations of family households rather than with populations of working individuals classified by workplace. They are therefore most appropriately organised on the territorial principle by place of residence. In many Western countries there exist permanent territorially based organisations for the conduct of demographic and social surveys, such as those of the Current Population Survey in the USA and of the Social Survey Division of the Office of Population Censuses and Surveys (formerly the National Social Survey) in the UK, which provide a nation-wide infrastructure for survey work - a corps of trained interviewers distributed throughout the country, a centre of technical expertise, computer facilities, up-to-date sampling frames and so on. A central aspect of the neglect of social statistics in the USSR and of the weakness of territorial administration is the continuing absence of any corresponding survey network, even though Soviet demographers have pressed for its creation for more than two decades (Section 7). This impedes the proper conduct of surveys except as small-scale local experiments.

We are aware of only two social sample surveys of more than local scope conducted on the territorial principle by TsSU - the surveys of old-age pensioners and of invalids in the RSFSR conducted in 1973 and 1974 respectively (Inshutina 1975)[3]. These surveys both used a sampling frame

designed by the Sampling Laboratory of the Scientific Research Institute of TsSU. Inshutina envisages further surveys of pensioners using the same frame; a new survey of pensioners is at present being planned (Dmitriyeva and Volkov 1983).

It is surprising that a society as highly organised as the USSR should lack a nation-wide system of residential records usable as a sampling frame, but this does seem to be the case. Two systems the possible use of which has been discussed in the literature but which have been found inadequate are (a) the records kept in the urban residential administration offices (ZhEKi), supplemented with the records of the village Soviets in rural areas, and (b) the records compiled in the course of the population censuses. It is not clear why consideration is not given to the use of electoral registers, so often resorted to as sampling frames in the West.

Flats or households cannot be directly selected from ZhEK records, which only list buildings and show the number of residents in each. There must be doubts regarding their quality, for Boyarskii and Venetskii (1978) state that their use "will require a certain improvement in record-keeping in the residential administration offices, a more rigorous observation of set procedure etc.". The poor quality of village Soviet records will become evident shortly. Population census records are generally regarded as reliable, but rapidly become out-of-date in the intervals between the approximately decennial censuses. Volkov (1971) contemplates combining their use with that of ZhEK lists of newly constructed buildings; this seems to be the best solution presently feasible. It remains possible to use the infrastructure created to carry out a population census for a simultaneous sample survey, an opportunity which was taken in 1970 and 1979 (Shenfield (1982a pp.22-24).

Reliable sampling frames may be compiled in particular areas in connection with special exercises. For example, the experimental budget

survey conducted in Riga in 1967-8 on the territorial principle by the method of "momentary observations" (and discussed in the next chapter) was able to rely on a sampling frame constructed three years earlier, in 1964, in connection with a once-off census of young people living in Riga. The Latvian TsSU had at that time compiled careful lists of all buildings by neighbourhood, indicating the number of flats in each. The availability of these lists was one of the reasons Riga was selected as the location of the experimental survey (Vitols 1972). By implication the non-existence of satisfactory records in most parts of the USSR at most times is confirmed.

Among the most serious consequences of the undeveloped state of territorial sampling is the inadequacy of population statistics in the periods between censuses. Data on fertility, migration and other topical issues are collected by research institutes, but their surveys are almost always on a very local scale and are often very poorly designed (Volkov 1971). Reliance must in general be had on "current population accounting" (tekushchii uchet naseleniya) - the statistics derived from certificates of birth, death, marriage and divorce from the administrative records kept by internal-passport offices and village Soviets. Shubina (1973), of the Department of Statistics of Population, Health and Social Security of TsSU USSR, has assessed the quality of these records by comparing statistics derived from them with data from the 1970 population census.

Shubina shows the incompleteness and inaccuracy of the population records kept by the personnel of village Soviets; these records are described by Dmitriyeva (1980). This must be partly attributable to the low level of education of these personnel (see Chapter C1). New arrivals to a village are often not registered in the records, while young people who leave the village "temporarily" to study but in fact fail to return remain in the records as residents, leading to an underestimation of

rural-to-urban migration in the statistics.

Thus in 1968 the rural population of the USSR was overestimated by over four million. Village records improved during 1968 as a result of the issue of new record books, and continued to improve during 1969 because of fear that negligence would be shown up when the records were compared with data from the forthcoming population census. The estimate of the rural population at the beginning of 1970 was accordingly "only" about one million too high. After the clean start provided by the census, village records showed a migration out of rural areas of 4.1 million over the three years 1970-72, while the much more reliable records kept by urban passport offices showed a migration into urban areas of 5.6 million over the same period - a new discrepancy of 1.5 million. However, Shubina concludes not that population statistics need to be reorganised on a sounder basis but merely that more effort must be put into checking the reliability of current population accounting.

Similarly, the Departmental statistics collected from medical establishments by the Ministry of Health are of limited relevance to any analysis of the health of the population (Semenova and Volkov 1959) and, moreover, are mostly of extremely poor quality, but proposals periodically advanced since the early 1950s to replace the bulk of these records by territorially based sample surveys have been only partly adopted (Shenfield 1982a pp.24-5).

In the absence of a social survey infrastructure organised on the territorial principle, the workplace-based infrastructure of the family budget survey is used by TsSU for those other social surveys which it occasionally conducts (see Appendix 1 to Chapter B1). Most of the social surveys conducted by social scientists outside TsSU are likewise workplace-based. According to Shlyapentokh (1976 pp.186-7), selection by place of work is used in at least two-thirds of all Soviet social surveys, and was

used in 90 per cent of the surveys conducted in the period 1970-73. Even of those surveys dealing with family problems bearing no relation to employment 50 per cent are workplace-based. For example, Antropova and Melnikova (1981) investigate the influence of living conditions and of such problems as family tension and alcoholism on the morbidity of workers at a particular enterprise. On the other hand, there are some well-designed surveys organised on the territorial principle - in particular, those conducted under the leadership of Academician Zaslavskaya in Novosibirsk oblast' on such topics as rural migration and the use of non-working time by the Institute of the Economics and Organisation of Industrial Production.

4 The territorial and the branch principles in the FBS

Although it is often stated that the FBS is organised on the branch principle, a review of how it is organised leads rather to the conclusion that the territorial-branch principle is dominant here as in most of the activity of TsSU. Interviewers attached to enterprises report directly to the oblast' statistical administrations, with only a minor supervisory role for the Inspectorates of State Statistics at raion and town level. Sample design and data processing within each oblast' follow the branch principle. At higher levels, however, the territorial principle plays a role alongside the branch principle - in the allocation of sample quotas (Chapter B1), in the lines of communication from oblast' through Republican to All-Soviet offices, and in the preparation of reports by branch for particular oblasti, economic regions and Republics. Even the biases in the sample are both of a territorial and of a branch nature (Chapters B3 and B4).

The use of the branch principle at oblast' level "has a number of organisational advantages" (Venetskii and Matyukha 1971) which derive from the domination of the branch principle at this level in Soviet administration and statistics. Apart from the ready availability of reliable sampling

frames of enterprises in the oblast', it provides for access to enterprise resources in support of survey work, resources which it would be much more difficult to extract from weak local authorities. We have already given various examples of the use of enterprise resources in Chapter B5: the enterprise bookkeeping office serves as a source of documented wages data: participation in the survey is supervised by enterprise management through Commissions for Assisting the Budget Survey and the like; enterprises are prevailed upon to provide transport to meetings of participants (Vsesoyuznye ... 1954); enterprises award bonuses to participants for conscientious record-keeping.

In the "microcensus", the periodic survey of incomes, living conditions and family composition, even more use seems to be made of enterprise resources. "Technical preparation" as well as interviewing work is imposed on the management and social organisations of enterprises (Rimashevskaya 1965). Enterprise staff are used as microcensus interviewers (Matyukha 1960, Vsesoyuznoe... 1969 p.256)[4]. An article by Berzkaln (1960), advocating the use of punched-card equipment belonging to large enterprises to mechanise the labour-intensive process of compiling ordered card-files of workers and employees from which to select samples for the microcensus, suggests that enterprise labour is also commandeered for sample selection. The organisation of the microcensus on the territorial-branch principle is connected to the similar organisation of the FBS, as recognition that the principle is unsatisfactory for the one would imply that it is unsatisfactory for the other too.

5 Arguments in criticism of the existing territorial-branch organisation of the FBS

The arguments brought by critics against the existing principle of organisation of the survey fall under four heads. First, the incomplete

and biased coverage of the survey is often attributed to the role of the branch principle. Second, it is argued that the allocation of families to branches is inevitably arbitrary and misleading. A third argument, linked to the second, is that the data breakdowns by branch facilitated by the organisation of the sample on branch lines are of little or no interest, while other and much more important breakdowns are thereby impeded. Finally, some statisticians have made the point that the sampling efficiency of a territorially based survey would be higher than that of a branch-based survey of equal size. We shall consider these arguments in turn, and then show how the critics support their case by reference to foreign experience.

Several of the biases to which the sample is subject are attributed by critics to use of the branch principle: the exclusion of branches in the "non-productive" sphere (eg. Karapetyan, Rimashevskaya and Sidlyarenko 1967), the exclusion of unemployed groups such as old-age pensioners and invalids (eg. Shlyapentokh 1976), the uneven territorial distribution of the sample (eg. Levin 1969 p.151) and the multi-worker bias (eg. Shvyrkov 1965, Ananyeva 1966, Safronova 1968). Reconstruction on the territorial principle is then alleged to be necessitated by the need to eliminate these biases.

The argument is, however, a little dubious. Reconstruction on the territorial principle would be the surest and most sensible way of solving these problems, but much can be done to counteract the biases even while continuing to use the branch principle. The coverage of the survey has been extended at one time or another to previously excluded branches and territories, and recently even old-age pensioners have been separately surveyed. The microcensus, also organised on the branch principle, covers all branches of the economy and all territories of the USSR, and there is no technical reason why the FBS could not do likewise. Two methods of

correcting for the multi-worker bias were mentioned in Chapter B1, and it is always open to defenders of the branch principle to allude to them in order to "show" that reconstruction of the sample is not necessary (for example, Venetskii and Matyukha 1971). The critics have a tendency to make the branch principle into a scapegoat for all the deficiencies of the survey design, or perhaps they merely wish to assemble as many arguments as possible for a reconstruction which, after all, has ample justification.

It is helpful to distinguish two aspects of the question of the arbitrary nature of allocations of families to branch categories: whether the least arbitrary allocations possible are being made, and whether even these are still so arbitrary as to deprive the categorisation of all value.

Families are most appropriately classified according to indicators of the position of the family as a whole. However, only individual workers and employees can be classified by branch; only families all working members of which work in the same branch of the economy can be allocated to branches without resort to an arbitrary convention which defines the branch of a family as the branch of one or other family member. Under existing Soviet procedure, the member who determines the branch of his family is simply the one who happens to be selected for the sample at his workplace, who may earn only a small proportion of the family income. The use of the "heads of family" method would ensure that it was usually the chief earner who determined the allocation of the family. This less arbitrary situation can be attained without abandoning the branch principle, so once again reference to the unsatisfactory nature of existing practice (eg. by Ananyeva 1966 and by Safronova 1968) does not qualify as proof of the necessity of territorial reorganisation.

Allocation of families to branches is inevitably more or less arbitrary to the extent that different members of the same families work in different branches. The use of the branch principle in pre-revolutionary

budget surveys could be justified by the fact that at that time members of the same family of workers tended to work in the same branch (Karapetyan 1980). This is much less frequently true nowadays; similarly, the number of families containing both workers and employees grows over time. Matyukha (1973 p.132) attempts to contest this point, claiming that "a certain arbitrariness in the classification (of families by branch) ... does not reduce the great practical significance (of the survey) for investigating living standards". He adduces the fact that in 1972 the average family of workers and employees had 1.7 working members, of which only 0.3 worked in a branch other than that in which the basic worker of the family worked. This means, however, that approximately 40 per cent (0.3/0.7) of non-basic workers worked in other branches, quite a substantial proportion.

Many critics argue that data breakdowns by branch of family, however defined, are of negligible interest anyway. Branch has an insignificant influence on the pattern of consumption (Shvyrkov 1965). Karapetyan (1980) asks:

> What sense and cognitive value can there be, for example, in a survey of the budgets of oil workers scattered over regions with completely different living conditions and structures of consumption (Azerbaijan, Tatar ASSR, Tyumen' etc.)?

In view of the concern with regional differentials in living standards, he argues, it is the territorial variable which is the most significant. As we have seen, the branch principle is blamed for the lack of territorial representativeness of the sample which impedes the production of reliable territorial data breakdowns. Furthermore, the branch principle divides the sample into workers and employees on the one hand and collective farmers on the other, a division seriously discrepant with that into the urban and the rural population, much more significant for the study of living

standards, because many workers and employees live in rural areas (Shvyrkov 1965, Levin 1973). The branch principle does not facilitate the production of important data breakdowns by family composition and by level of per-capita income (Shvyrkov 1965), though nor does it make them technically impossible (Chapter C3).

The views of A.I.Parfenova, Head of the Budget Survey Department of TsSU RSFSR, on how to design the sample so as to facilitate relevant data breakdowns are of great interest, both on account of their originality and because she is the only TsSU official on record as recognising that "the branch principle contributes little to (the production of) all-sided and fully representative data" (Sukhoruchkina 1970). Analysis of budget data for workers and employees in Moscow oblast' by TsSU RSFSR, she reports, shows that consumption depends neither on place of residence nor on branch of employment, but on per-capita income and family composition. (In view of the geographical limitation on the data analysed, one may think her conclusion that place of residence is irrelevant a little hasty!) Sample selection should therefore follow neither branch nor territorial principles, but should be based on family types - single-person households, families consisting of two workers, of one worker and one dependent, different types of three-person families and so on. Not only the number but also the age of children must be taken into account in this stratification, as both the private and the social consumption of families depend on the age of their children. Unfortunately we have no further information on this scheme or its fate. The difficulty of constructing suitable sampling frames for it casts doubt upon its practicability.

The final criticism made of the branch principle is of a more technical nature. As we pointed out in Chapter B1, cluster sampling, necessary for administrative reasons, increases sampling error for any given sample size relative to non-cluster sampling using the same system of

stratification, thereby reducing sampling efficiency measured in terms of precision per sampling unit. This is because the sampling units comprising a cluster do not in practice have independent characteristics; their indicators are more or less highly correlated. The loss of efficiency depends on the degree of correlation and therefore on the type of clustering used, because clusters formed in different ways have different intra-cluster correlations. Ananyeva (1966) and Safronova (1968) state that intra-cluster correlation among families living in the same neighbourhoods (population points) is significantly less than that among families of workers in the same enterprises. Thus it is more efficient to select from residential than from workplace clusters. One factor here may be that Soviet neighbourhoods are fairly mixed in their social group composition, by comparison with neighbourhoods in Western countries. Organisation of the survey on the territorial principle would therefore enable greater precision to be attained with the same sample size, or the same precision with a smaller sample size.

One tactic used by critics of the Soviet budget survey to publicise their views is the publication of reports of budget survey methods in foreign countries where the survey is designed in a sound fashion and on a territorial basis. (These conditions exclude the surveys in many countries in the West - for example, the FRG - as well as in the East.) Sometimes, as in the admiring report of survey sampling practice in the USA by Semenova and Volkov (1959), conclusions are explicitly drawn concerning how the methods being reported "can also be used for various purposes to great advantage in Soviet conditions". More often this conclusion remains implicit, as in the description of the Indian budget survey by Zhutovskaya (1958), but one easily drawn by the knowledgeable reader.

As might be expected, the most common point of reference abroad is the non-Soviet countries of Eastern Europe. Three reviews of Eastern

European practice have been published in the Soviet literature - those by Ananyeva (1966) and by Kazun (1976), and a report by Venetskii and Matyukha (1975) of a meeting of Eastern European statisticians to discuss budget survey practice. Most of the Eastern European countries follow the Soviet model and conduct their surveys on the branch principle; this is true of the GDR, Poland[5], Czechoslovakia[6] and Rumania. The post-war budget survey in Hungary has been conducted on the territorial principle since it was set up in 1949, but cannot serve as a model for Soviet critics because, like many Western market research surveys, it uses quota sampling, a method which lacks a sound foundation in sampling theory[7]. Soviet commentators prefer to cite the example of the Bulgarian budget survey. Not only is the current sample design sound and territorially based, but the survey has undergone the sort of reconstruction advocated by Soviet reformers for their own survey. From the time it was set up in 1953 until 1961, the post-war survey in Bulgaria was also organised on the branch principle, but was reconstructed on a territorial basis in 1962, as reported by an official of the Bulgarian Central Statistical Office (Tsanov 1963) in the journal of the Soviet TsSU[8].

6 Arguments in defence of the existing territorial-branch organisation of the FBS

TsSU officials generally ignore the criticisms of the existing organisation of the FBS made by those advocating the reconstruction of the survey on the territorial principle. However, three basic arguments have been put forward by them in defence of the existing survey: that the bias of the existing sample is not in fact very great, that it is not necessary for the FBS to represent the whole population, and that bias can be readily corrected by recalculating the data. There is of course hardly any logical consistency among these three arguments.

An example of an argument aiming to minimise the bias of the existing sample was given in the last Section - that is, the effort by Matyukha (1973) to understate the multi-worker bias.

For an example of an argument denying the need for the FBS to represent the whole population, we may cite Postnikov (1961), who explains why FBS data do not describe the consumption of the population of Tadzhikistan as a whole and goes on to comment:

> And this is not necessary: the consumption of the population as a whole can be calculated from production and distribution data by the balance method.

This point is self-contradictory, inasmuch as the balances to which reference is here made themselves rely to some extent on FBS data (see Chapter D2).

In the third argument the use of corrective coefficients to adjust FBS data is presented as a simple and convenient alternative to reconstruction of the survey (see Chapter B9). The difficulties of deriving corrective coefficients are understated and the effectiveness of their use overstated.

In references to the "difficulties" which reconstruction of the survey on new lines would involve, we can sense the resistance of officials whose main desire is for an easy life to the prospect of change which would disrupt familiar routines. In conversation with the writer, Rimashevskaya commented:

> TsSU is the most conservative State body. It is something you can feel in your bones. They are more resistant to change than Gosplan, for example, who have a direct responsibility for achieving coordination in planning and so are more prepared to try out various methods. TsSU are used to working within a rigid system which change would disrupt. "The game isn't worth the candle", as the saying goes (Shenfield 1982b).

One element in the resistance to change is the enormous importance attached to the calculation of year-on-year trends, so that reports can

take the customary format: "The consumption of product X is Y per cent up on the corresponding period of last year". This exaggerated concern with percentage changes over time, to the neglect of absolute levels, is typical of Soviet statistical work. Changes of methodology make it difficult or even impossible to calculate such percentage trends over periods which span the change, and this is treated as a serious obstacle to change.

A common view used to support giving higher priority to the maintenance of comparability over time than to the improvement of data reliability by reforms in methodology is that, even if existing methods produce unreliable data on absolute levels, they at least produce tolerably reliable data on trends over time. That such an argument should be found convincing is in itself suggestive of the exaggerated importance attached to trend data. The unjustified assumption that one need not worry about trend data can also be found in the writings of Soviet economists outside TsSU (and, indeed, in those of Western scholars as well). For example, Rogova and Rakovskaya (1981) assure us that:

> Although insufficiently representative in their absolute levels, budget survey data on monetary resources by income groups may be used to reveal basic trends over time...

Shvyrkov (1965) provides us with an example of how bias in survey data can falsify trends over time as well. The rate of growth of money incomes of families of workers and employees in the RSFSR over the period 1958-61 was underestimated by the budget survey by about two-thirds of the true rate, as estimated by the balance of money incomes and expenditures of the population. This was the result of the change for the better in the position of the worst-off families at this time, when welfare policies in their favour (minimum wages, family allowances etc.) were being introduced. The change could not be properly reflected in budget survey data because

the low-income families affected were (and are) drastically under-represented in the survey sample.

7 The proposals for reconstructing the FBS on the territorial principle

Economists writing on questions of welfare often mention that the budget survey should be reorganised on a territorial basis, but for more detailed expositions of this proposal we must turn to the work of a number of mathematically sophisticated demographers, and in particular to that of A.G.Volkov, whose name has been most consistently associated with the issue for over two decades. For the demographers, however, reconstruction of the budget survey is not an isolated problem but only one aspect of the creation of an integrated territorial network for multi-purpose sample surveys of the population.

This project seems to have been first advocated by Semenova and Volkov in 1959, who described the US Current Population Survey, arguing that the USSR had much to learn from its methodology. TsSU should "give up stagnation and tradition" and "take account of foreign experience" by setting up a similar "sample network for observation of the population" for the flexible study of demographic, health and welfare matters, including those covered by the budget survey, which was to be superseded by the new framework. It was even suggested that oblast' statistical administrations might conduct surveys of local interest on their own initiative ("of course, with the agreement of the centre"). The 1959 population census would provide a basis for constructing the network: "we must not let this opportunity slip by". In the event, the opportunity presented by each of the censuses of 1959, 1970 and 1979 has been let slip by, though a sampling component has been incorporated into the last two censuses.

Since 1959 the project for a survey network has been more thoroughly worked out and has gained a great deal of support (Table B10.2). A series

TABLE B10.2

DATES IN THE DEVELOPMENT OF THE PROPOSAL FOR A TERRITORIAL SOCIAL SURVEY NETWORK

1959 Semenova and Volkov urge that the 1959 population census be used to create a survey network similar to the US Current Population Survey.

Proposals for the use of sampling in the population census rejected.

1962 Volkov reports research work in progress for a survey network based on records of the residential administrations.

1966 Volkov argues the case for a survey network at a meeting of a Commission of the Scientific-Methodological Council of TsSU USSR.

1968 Safronova submits her critical kandidat dissertation on "sampling in the practice of Soviet statistics" to the Moscow Economic-Statistical Institute (MESI).

Pavlova, of the Scientific Research Institute of TsSU, advocates a survey network at the symposium "Soviet statistics in twenty years' time" (Simpozium... 1968).

1970 Sampling used in the population census for the first time.

Orekhova submits her dissertation on "specially organised observation" to the Leningrad Financial-Economic Institute, advocating learning from the multi-purpose social surveys of the 1920s.

1971 Volkov submits his dissertation on "sample observation of the population", prepared in the Scientific Research Institute of TsSU, to MESI. It includes the first thoroughly worked out scheme for a survey network, based on the population census.

1972 Paper by Inshutina "on the possibility of applying multi-purpose sampling in social statistics" published by MESI.

1973 Sampling Laboratory of the Scientific Research Institute of TsSU conducts a territorial sample survey of pensioners in the RSFSR.

1974 Sampling Laboratory of NII TsSU conducts a similar survey of invalids in the RSFSR.

Aleksandrov submits his dissertation on "stratified multistage and multiphase samples and perspectives for their application" to Moscow University Economics Faculty.

1975 Inshutina submits her dissertation on "application of multi-purpose sample surveys in social statistics" to NII TsSU.

1976 Article on "tasks of population statistics and demographic research in the tenth Five Year Plan" published under the names of Dmitriyeva (Department of Statistics of Population, Health and Social Security of TsSU USSR), Boyarskii and Volkov (NII TsSU). The idea of a survey network is raised and tentatively approved.

1978 Detailed programme for a survey network published under the names of Boyarskii and Venetskii. Mentioned as having been discussed and approved at a session of the Scientific Council of NII TsSU USSR.

of kandidat dissertations have been partly or wholly devoted to the problems involved (Safronova 1968, Orekhova 1970, Volkov 1971, Aleksandrov 1974, Inshutina 1975). Volkov's own dissertation included a well worked-out design for a survey network, based on analysis of data from the 1967 trial population census carried out by the Laboratory of Demography of the Scientific Research Institute of TsSU USSR (NII TsSU). In this way researchers in at least three centres - the Moscow Economic-Statistical Institute (MESI), Moscow University Economics Faculty and NII TsSU - have become familiar with and supporters of the project.

In 1976 an article by Dmitriyeva, Head of the Department of Population, Health and Social Security Statistics of TsSU USSR, Boyarskii, Head of NII TsSU, and Volkov, Head of the Department of Demography of NII TsSU, setting "the tasks of population statistics" in the coming tenth Five Year Plan, appeared in Vestnik statistiki. An article of this kind co-authored by an official of TsSU itself and not just by representatives of the TsSU research institute, has a certain official status. Its account of the advantages of a survey network represents the closest point the proposal has ever come to being officially accepted by TsSU, though the language used is far from that of unequivocal advocacy: "we must discuss a significantly broadened application of sampling... This leads to the question of the creation of a permanent sample network for observation of the population..." In 1978 Boyarskii, together with Venetskii (of MESI), had published in Vestnik statistiki the most substantive programme yet for the survey network, based on research in NII TsSU and labelled as having been discussed and approved at a session of the Scientific Council of NII TsSU. An academic consensus had been established but there had been no breakthrough in TsSU's position.

In fact, a new article by Dmitriyeva and Volkov (1983) on "the tasks of population statistics" marks a definite setback for the scheme. No

mention at all is made of the very possibility of a survey network. Volkov must surely remain convinced of the value of the idea to which he has devoted so much work, but is presumably less able to moderate Dmitriyeva's resistance to it than in the past. Volkov no longer enjoys the support of the Head of NII TsSU, Boyarskii having been replaced in this position by Eidelman, an input-output specialist who evidently lacks his predecessor's interest in socio-demographic surveys.

The article reports work in progress since 1979 to strengthen population statistics by an alternative strategy which may have been chosen in explicit preference to the survey network. The strategy concentrates on improving the data on which Dmitriyeva's Department traditionally relies, those derived from the records of current population accounting. New questions have been added to the records and yet more may still be added. Processing of the data is to be computerised and considerably expanded. The data collected have been made consistent with the data from the 1979 population census so that the two sources can be used jointly: rates of demographic processes (natality, mortality etc.) are to be calculated by dividing absolute numbers of demographic events (births, deaths etc.) taken from population accounting by corresponding base populations taken from the census.

Based on complete enumeration rather than on sampling, this is an extremely expensive and inefficient strategy. Nor is it likely to yield very reliable information: it is admitted that population accounting records remain of poor quality, while census data go rapidly out of date. The investment of resources in the population accounting system will make it all the more difficult to cut losses and decide to establish a survey network instead.

Although in 1962 Volkov, presumably reluctant to wait for the next census, suggested basing the survey network on ZhEK records, in his 1971 dissertation he came down in favour of using the more reliable frame of a

population census as a starting point, and this seems to have become a consensus view of the reformers. It implies that the chance of establishing the network comes only every ten years, so that the earliest possible date is now about 1990. Boyarskii and Venetskii (1978) describe four possible types of sample structure which could have been based on the 1979 census, and report an experimental ten per cent survey conducted by NII TsSU in two raiony, one in Latvia and one in the Ukraine, based on the record booklets of the trial census. The questionnaire covered a sample of the topics which they proposed should be researched by the future survey organisation (Table B10.3) - general socio-demographic description of family members, intentions to purchase consumer durables, and attitudes to radio and television programmes.

Volkov (1971) proposed that the survey work to be carried out on the network comprise a stable core survey with a large sample but a short programme restricted to basic demographic data, and a flexible series of surveys on sub-samples to collect other data, including family budgets, and to study topical problems as they arise, either on a local or on a nation-wide scale. His calculations show that the core sample would have to be of the order of 500,000 people, or 0.2 per cent of the population, to give regional demographic indicators of the necessary precision. He outlines a four-stage stratified sample design with the following sampling units at each stage:

First stage: oblast'-level territorial units, stratified on demographic and economic criteria, with two oblasti selected from each stratum but the largest units included obligatorily;

Second stage: raiony, with separate urban and rural strata;

TABLE B10.3

LIST OF SURVEYS PROPOSED FOR THE SAMPLE NETWORK BY BOYARSKII AND VENETSKII (1978)

I. Reproduction and structure of the population

Natality, mortality, size and composition of family, marriage and fertility, family planning, divorce, the demographic role of ethnic factors etc.,

II. Labour resources

Training of youth, occupational attitudes of youth, use of the labour resources of small towns, use of women's labour in the national economy etc.,

III. Social hygiene

Socio-economic influences on mortality, social-hygienic assessment of living conditions, morbidity, occupational and other injuries

IV. Migration

Characteristics of migrants, urban-rural differences, demographic problems of towns and settlements with a narrow industrial specialisation etc.,

V. Living standards

Study of incomes, expenditures and consumption; study of living conditions; requirements for everyday services, social and medical services; time budget surveys

VI. Public opinion surveys

Third stage: instructors' plots (instruktorskie uchastki) or the smaller enumerators' plots (schetnye uchastki) of the last population census;

Fourth stage: residential addresses (zhilye pomeshcheniya), with coverage of all persons and families living at the selected addresses, and separate coverage of hostels.

The second and third-stage units could be selected with probability proportional to population. To avoid the practical inconvenience of an extremely dispersed sample, at least forty addresses should be selected in each raion, but clustering at levels more local than this is rejected because too much efficiency would be lost as a result of intra-cluster correlation. Volkov envisages that the work will be facilitated in the future through the creation of an automated register of the population.

8 Factors impeding reorganisation of the FBS on the territorial principle

There is a long tradition behind the use of the branch principle in budget surveys in the USSR; the very first budget surveys in pre-revolutionary Russia were organised on an industrial basis (Table B10.4). That the branch principle should have retained its relevance during the period of Soviet industrialisation is explained by Karapetyan (1980 p.244) by the primacy at that time of monitoring the welfare of the workforce of the highest-priority, mainly heavy-industrial, branches. Exclusive concern with the welfare of a key elite of workers would also explain the exclusion of the unskilled from the budget survey. Now that the USSR has reached a stage of development at which it can raise the living standard of the whole people, irrespective of branch, and above all of the badly-off strata, argues Karapetyan, this factor has lost its significance, but the force of inertia holds back the appropriate adjustments in statistics.

We may wonder whether a residual indifference to the welfare of the

TABLE B10.4

PRE-REVOLUTIONARY BUDGET SURVEYS OF WORKERS' FAMILIES

Date	Main researcher	Population under investigation	Number of budgets collected	Number of budgets processed
1907-8	S N Prokopovich	Petersburg factory workers	1016	570
1908	M N Davidovich	Petersburg textile workers	50	41
1909	I M Shaposhnikov	Textile workers in the town of Bogorod (near Moscow)	340	324
1910	A M Stopani	Oil workers in Baku (Azerbaijan)	2339	2244
1911	V P Goritskii	Textile workers in Sereda (Kostroma guberniya)	20	18
1913	G Naumov	Handicraft and factory workers in Kiev (Ukraine)	750	572

Source: Matyukha (1967 pp 174-5)

worst-off strata may not still play some role, but we suspect that the main impediment to change is the resistance of vested organisational interests. Reconstruction of the budget survey on the territorial principle is not an isolated proposal which could be implemented by the existing TsSU Department of Budget Statistics, even assuming it had the will and the expertise to do so. It is, as we have seen, just one element in a wide-ranging programme to put the whole of Soviet social and demographic statistics on a sounder basis. The advocates of this programme are usually silent in their writing on the question of the organisational form it might take, but Inshutina (1975) suggests that a special statistical office be set up, presumably outside TsSU, to plan, coordinate and conduct surveys on a khozraschet basis - that is, with the commissioning and funding of surveys by client organisations. If the survey network were to be set up by TsSU, a new Department would have to be created for the purpose, for the wide-ranging field of research envisaged would extend far beyond the sphere of competence of any of the existing specialised TsSU Departments; in particular, the Department of Budget Statistics would inevitably disappear.

Those most closely associated with the project for a social survey network, such as Volkov and Inshutina, are now mainly based in the Sampling Laboratory and the Demographic Laboratory of NII TsSU. It is likely that they would constitute the core group of any new survey organisation created within or outside TsSU. Other pools of expertise in the field of social and demographic survey work, such as Zaslavskaya's group in Novosibirsk, might also be drawn upon.

Be that as it may, there is no doubt that creation of a survey network on the lines envisaged would entail a substantial rearrangement of bureaucratic interests and a shift in the orientation of the Soviet statistical system away from the near-monopoly of attention enjoyed by its

traditionally dominant economic concerns. It is hard to see how this could happen other than by high-level intervention in favour of the reformers, involving perhaps the appointment of reform-minded specialists to leading positions within TsSU. Such a change would be welcomed by a broad constituency of Soviet scholars interested in more reliable and relevant statistics in the fields of demography, public health, market research, public opinion, sociology and labour economics as well as in family budget research.

9 Conclusions

The parts played by the territorial and branch principles in the organisation of the FBS are consistent with the parts played by them in Soviet statistics and administration generally. Although those who advocate the reconstruction of the survey on a fully territorial basis have an overwhelming technical and social-scientific case, their proposals go against the grain of Soviet life, and this may sum up the reasons they have had no success to date.

Notes to Chapter B10

1 There have recently been efforts to strengthen "local" controls outside the biggest cities (Fuller 1983). It should be noted that the local Party apparatus enjoys more effective power than the local Soviet authorities.

2 "Local" industry, under Republican Ministries of Local Industry, is a partial exception to the dominance of the branch principle in industry.

3 The Central Institute of Expertise on Labour and Invalidism took part in organising these surveys.

There may of course be nation-wide territorially based surveys conducted outside TsSU which have escaped our notice. The survey of rural conditions by Arutyunyan (1971) amounts to a collection of a few local case-studies rather than a nation-wide survey.

4 In the 1958 microcensus, "social organisations" (Party, Trade Union etc.) at the enterprise carried out "explanatory work", and enterprises allocated about 5000 clerical personnel for training by the statistical administrations as survey interviewers (Matyukha 1960).

5 A note appears in the Express-Information column of Vestnik statistiki (1972, 7, 82) about a report by Kordos (1971) of two experimental studies in Poland in 1968-9 in which budgets were collected using a rotating sample and periods of participation of 2-4 weeks. One of these surveys was based on place of residence, the other on place of work. Lectures were delivered on the use of rotating samples in Poland and Bulgaria at the international meeting in 1975. A footnote by Rimashevskaya in Karapetyan (1980 p.31) states that Poland as well as Bulgaria has now made the transition to use of the territorial principle, and this is consistent with the fact that Kordos has been in charge of the Polish budget survey in recent years.

6 However, the microcensus in Czechoslovakia is organised on the territorial principle (Matyukha 1963).

7 The Hungarian budget survey also plays a role in planning different from that of the Soviet survey in consequence of the different economic mechanisms operating in the two countries. For example, in Hungary - as in Western countries and in contrast to the USSR - the survey is used as a base for the construction of retail price indices.

8 Bulgaria was divided into 6500 plots of about 600 families each, with large population points divided into two or more plots and very small population points combined to form plots. A stratified sample of 314 plots was selected, with proportional representation by district (okrug) and of urban and rural plots within districts. Plots were selected systematically from district lists, and within each plot eight families were systematically selected for the survey, giving a total sample size of about 2500. Intra-cluster correlation being significant, it was decided to reduce cluster size from 25 (in the workplace clusters before reorganisation) to 8; it was estimated that this reduced the sampling error of estimates of average per-capita income by 59 per cent relative to territorial clusters of size 25. These calculations confirm that those who redesigned the Bulgarian survey were familiar with mathematical statistics.

CHAPTER B11

THE METHOD OF MOMENTARY OBSERVATIONS

1 Introduction

In the last chapter we considered the criticism most commonly made of the sample design of the Soviet family budget survey, its use of the branch principle, and reviewed the proposals for correcting this deficiency by reconstructing the survey on the territorial principle. In Chapter B5 we discussed the harmful consequences for the representativeness of survey data of another serious deficiency in the sample design, the excessive length of the periods of time that families participate in the survey. In this chapter we review the main proposal made in the Soviet literature for reducing periods of participation - the "method of momentary (or sudden) observations" (metod momentnykh (vnezapnykh) nablyudenii).

The method of momentary observations was developed as a work-study technique to monitor the activity of workers and of productive equipment both in the USSR and in the West, where it is known as "activity sampling". Soviet researchers have adapted the technique for use in budget surveys - a technological transfer which, so far as we are aware, has not been pursued in the West. We shall therefore be able to compare the problems of introducing the same technique in two different fields of application.

In Section 2 we provide a general explanation of the method of momentary observations. Sections 3 and 4 deal with its development and use in industrial work-study and in budget surveys respectively. In Sections 5 and 6 we assess the main experimental exercise by which the method was made into a practical tool of budget study, the budget survey conducted in Riga, Latvia, in 1967-8, concentrating on the sample design in Section 5

and on the data analysis and the findings in Section 6. In Section 7 we discuss the implications of this issue for methodological innovation in Soviet statistics.

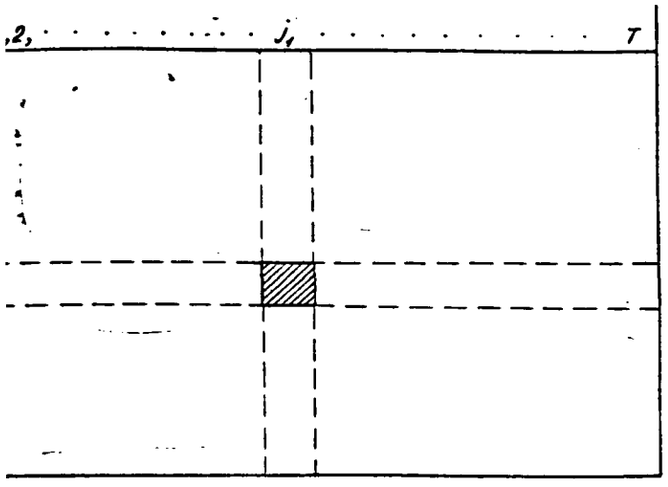
2 General explanation of the method of momentary observations

A heuristic description of the method of momentary observations will be sufficient for our purposes. The reader interested in the mathematical theory may refer to Davidson et al (1960) for an English-language analysis of the sampling error in activity sampling, and to Volkonskii (1962) for a derivation of the optimal parameters in the budget survey application.

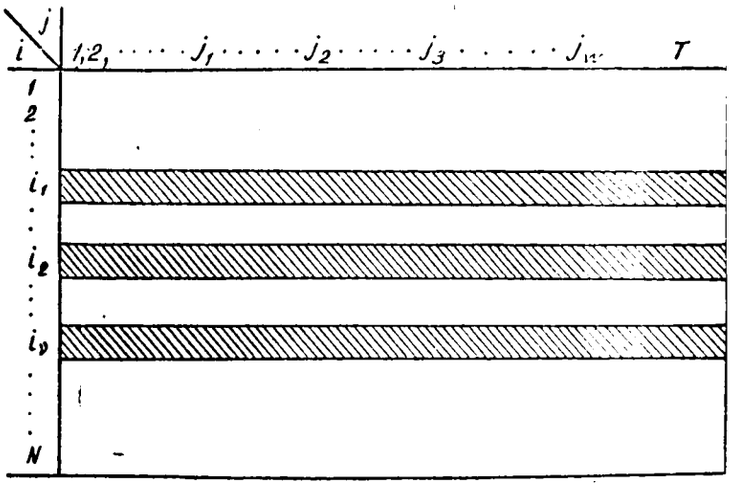
In Table B10.1 we reproduce from a Soviet account of the theory of the method (Balansy... 1969 pp.54-66) geometrical representations of different methods of sampling in time and space. The space dimension represents the population of objects the activity of which is to be observed, whether workers or machines in a factory or family households. In a continuous survey (b), a stable set of objects is observed through time without interruption. Thus, a small number of machines may be observed throughout the whole of a working day ("photography of the working day"); the budget survey of TsSU is also ideally of this form. In a discontinuous survey (d), different objects are observed at successive "moments" of time with a return to the first object only after the whole population has been covered. The "moments" may be instantaneous, as when machine-tools in a workshop are observed at one-minute intervals as "working" or "idle" and the proportion of "working" observations is taken as an estimate of the level of activity of the workshop. Or the "moments" may be short periods of time, as when families are questioned about their purchases in the two days (say) before the interview. More than one family may be interviewed per day, but if the population is at all sizeable no family is interviewed more than once. Observations must be evenly distributed in time, so that results

TABLE B11.1

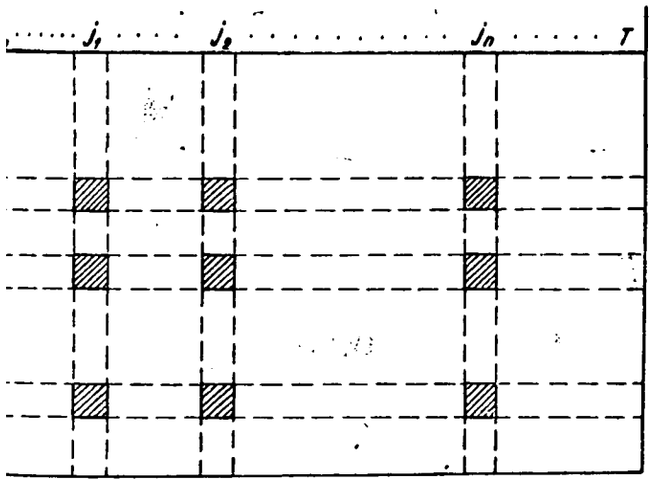
DIAGRAMMATIC REPRESENTATIONS OF DIFFERENT METHODS OF CONDUCTING BUDGET SURVEYS IN TIME



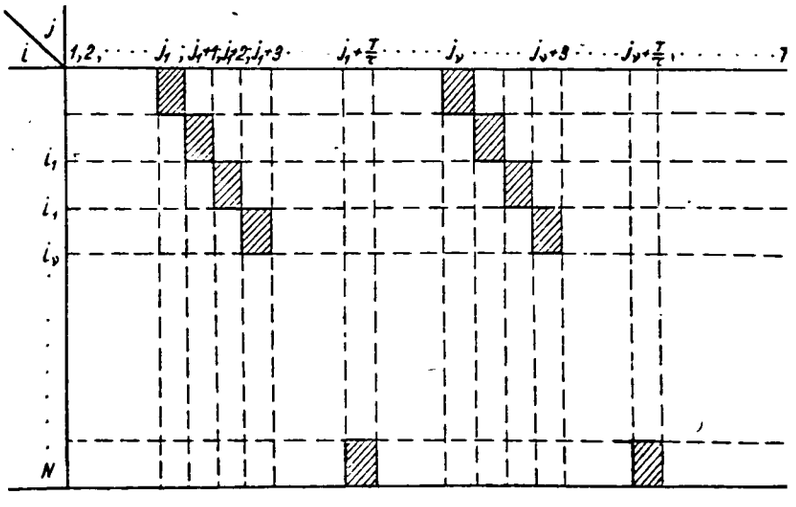
(a) General field of observation



(b) Field of continuous budget surveys



(c) Field of short-period budget observations



(d) Field of observation of family budgets by the momentary method

j = 1, ... T short periods ("moments") of time

i = 1, ... N population of objects (families)

Source: Balansy... (1969, Figs. 6-9, pp 61-65)

are not affected by periodic fluctuations in budgetary behaviour.

A balanced budget of all incomes and expenditures cannot of course be compiled on the basis of information about the budgetary behaviour of a family over a very short time period. In practice it would be necessary to divide budgetary indicators into several groups according to their patterns of fluctuation over time, to determine the optimal length of "moment" for each group and to conduct a separate sub-survey to collect data for each group (Section 6). Thus individual family budgets are to be abandoned, and balance of incomes and expenditures achieved only at the level of population groups - an idea that Karapetyan (1962) admits will seem strange, because of force of habit, to Soviet budget statisticians.

Budget surveys using the method of momentary observations have two main attractions for their advocates. First, by eliminating the record-keeping burden on participants the reliability of data would be enhanced. In experimental surveys very little "negative reaction" on the part of participants has been experienced, and refusal rates have not exceeded one per cent (Vagradyan 1962). Second, the method allows much larger samples of families to be covered with given resources, making it easier to ensure that samples are representative. A much greater efficiency of resource use is achieved: a great deal of information is obtained when a family is observed for the first time, but thereafter the amount of information obtained soon starts to rise much more slowly than cost with increasing period of participation, which points to the rationality of a high rate of turnover of the sample (Karapetyan 1962).

3 The method of momentary observations in industrial work-study

A dominant concern of Soviet work-study literature, and the one on which we focus here, is the methodology of assessing unplanned work stoppages and machinery breakdowns during working time, or "intra-shift

losses of working time". It is established procedure that all stoppages should be recorded on "stoppage forms" (prostoinye listki) or in special journals (Zapselskii 1982). These records, however, are never regarded as being in the least reliable; Chernyavskii (1972 p.14) states that the stoppage forms underestimate time losses by tens of times. TsSU, which itself collects these notionally complete records, recognises their uselessness by conducting annual sample surveys "to clarify the real state of affairs" (Chernyavskii 1972). These surveys are carried out with enterprise staff as the observers, who usually choose to make the observations at the end of October, an especially busy pre-holiday period. They also use the inefficient method of "photography of the working day", in which a small number of machines are observed continuously throughout the day.

Onopriyenko (1968) tells us that the method of momentary observations was first experimentally used in Soviet industry as early as 1924 and he lists publications on the method dated 1931 and 1933. At that time, in the field of sampling in general, the USSR was at least as advanced as the West and probably more advanced. By the end of the 1930s work on the "scientific organisation of labour" had practically been curtailed in the USSR, and the application of sampling in statistics had also been seriously restricted. It was then in the West that activity sampling was developed and introduced into wide industrial practice. Only in the late 1950s and early 1960s were Soviet mathematical statisticians able to revive interest in the subject by publishing studies of foreign experience (French, American, Polish etc.) and by carrying out small-scale experimental surveys in various industries.

We must take special note of one of the participants in this revival, A.Kh.Karapetyan (1908-1977); for biographical data see Aganbegyan (1980). Karapetyan graduated from the mechanics-mathematical faculty of Moscow State

University in 1932, having specialised in mathematical statistics. For many years he taught and did research at Yerevan State University in his native Armenia, and then at the kafedra of statistics of the Moscow Financial Institute, where he was until 1960. His main interest during this period was the statistical study of the use of productive capacity, on which subject he defended his kandidat dissertation in 1962. At the same time he was already taking an interest in the study of living standards, on which he published a book in 1959. Thus he was in a position to apply his knowledge of mathematical statistics to both these fields, with the method of momentary observations occupying a central place in his work.

The first post-war practical experiment in work-study by the method of momentary observations seems to have been organised in 1963 by Fedotov (1966) at a large railway station, where members of the station staff recorded the state of railway wagons at one-hour intervals. Reports of work in other branches of the economy followed in succeeding years. Thus Balan (1972) cites data from surveys by momentary observation in two construction organisations, while Trofimova (1973) of Tomsk State University demonstrates the advantages of momentary observation over "photography of the working day" by using both methods in a section of an industrial enterprise with ten machine-tools and comparing the results.

Other work is assessed by Marina Bauman (1970) in her dissertation on the method of momentary observations, prepared in the statistical computing institute VGPTI TsSU under the supervision of the Computer Centre of the Armenian Gosplan, directed at that time by Karapetyan. She describes two momentary-observation surveys which NII TsSU had carried out in a Moscow instrument-building factory - a study of the use of working time of lathe operators in the mechanical workshop and a study of the time budgets of diploma-holding engineers. Three methods were used in the engineers' study and the results compared: continuous observation, momentary observation and

questionnaires. Momentary observation was found more efficient than continuous observation, and more reliable than the use of questionnaires, which led to under-estimation of time spent on non-engineering tasks. Bauman comments that surveys by momentary observation are not usually well based in sampling theory and sampling errors are rarely calculated; there is need for a model methodology.

Nevertheless, although a significant number of studies have used the method of momentary observations, "photography of the working day" remains a more widely known method. Ilyenkova (1966) describes its use in the fur industry, and Balan (1972) reports that it is the most common method used in construction. The work-study monograph by Chernyavskii (1972) is devoted exclusively to "photography of the working day" and does not even mention the method of momentary observations.

Zaurov (1966) and Fedotov (1966) report that the surveys of the use of basic equipment in the engineering industry frequently conducted by TsSU are based on uninterrupted observation of the selected machines. In 1962 an experiment was permitted in which the method of momentary observations was used alongside the usual method at several enterprises. Results from the two methods were compatible, but procedures remained unchanged in succeeding years. In 1964 TsSU organised observation of the use of productive equipment and of workers' time at all sovnarkhoz engineering enterprises throughout the day of September 21. Engineering and other qualified staff were employed on this task, as a result of which all stoppages of more than five minutes were recorded. Fedotov points out that the use of sampling, with momentary observations at five-minute intervals, could have achieved the same result with the employment of 2-3 per cent of the number of registrators actually used. He also suggests that the method of momentary observation might be used in the compilation of the statistical reports which enterprises are required to submit on their use of productive

equipment; a recent textbook for TsSU staff (Posobie... 1980 pp.314-6) gives no indication of such a change in practice. Thus TsSU steadfastly resists all efforts by specialists, including those of its own institutes, to improve its working methods.

4 The method of momentary observations in family budget surveys

Rimashevskaya confirms that the method of momentary observations was transferred from the industrial field to that of budget surveys as a result of the personal initiative of Karapetyan, whom she regards as her "teacher" (Shenfield 1982b). She recalls working with him on the project in Moscow at the Scientific Research Institute of Labour (NII-truda). According to Aganbegyan (1980), however, Karapetyan returned from Moscow to Yerevan in 1960 to head the new Laboratory of Economic-Statistical Research at the Institute of Economics of the Armenian Academy of Sciences and remained in Yerevan until his death in 1977. His laboratory did research on living standards in close collaboration with NII-truda, and later with the Central Economic-Mathematical Institute (TsEMI), to which Rimashevskaya and her colleagues were transferred, so that there is no doubt Karapetyan and Rimashevskaya did work together in some sense. In 1968 Karapetyan defended a doctoral dissertation on "statistical methods of study of the incomes and consumption of the population of the USSR". He founded and directed the Computer Centre of the Armenian Gosplan (State Planning Commission), where he was responsible for working out the subsystem "standard of living" of the automated system of Republican planning (ASPR). It is an indication of the conservatism of TsSU as an institution that Karapetyan, an academic statistician, was offered an official position not in the TsSU system but in the planning apparatus. From 1972 he did at least train staff for TsSU as head of the kafedra of statistics at Yerevan State University.

In the early 1960s Karapetyan was the scientific leader of work on

the application of the method of momentary observations to budget surveys carried out jointly by his Laboratory of Economic-Statistical Research in Yerevan and by the Department for the Study of Living Standards of NII-*truda* in Moscow. The Scientific Research Economic Institute (NIEI) of the State Economic Council (*Gosekonomsovet*)[1] also participated in the research (Vagradyan 1962). In November 1961 the two main institutes involved held the Second All-Union Scientific Seminar on Questions of Applying Mathematical and Statistical Methods to the Study of Consumption (*Primenenie*... 1962) in Yerevan. The list of institutions represented at the seminar included several research institutes and computer centres, the Department of Commodity Turnover of Gosplan USSR, and the statistical administrations of Armenia and the RSFSR, but not the Budget Statistics Department of TsSU USSR. A number of papers on the conduct of budget surveys by the method of momentary observations were presented (Karapetyan 1962, Vagradyan 1962, Volkonskii 1962). Volkov, at that time at NII-*truda*, presented a paper on demographic sampling in general.

Of the three papers on budget surveys, that by Karapetyan was a general discussion of the potential advantages of the method of momentary observations, that by Volkonskii (from Leningrad) was devoted to the mathematical theory of the method, and that by Vagradyan reported a preliminary assessment of the efficiency of the method based on analysis of 200 consumption diaries of families of Leningrad workers for the month of September 1956, taken from a budget survey organised by the trade unions (VTsSPS). Total monthly expenditures on nine important products by individual families ("through sums") were calculated together with "diagonal sums", obtained by adding the expenditures of different families for successive days, representing estimates of monthly expenditure which might have been made had the method of momentary observations been used. The variance of diagonal sums was only one-eighth of the variance of through

sums, implying that momentary observation would give the same precision as continuous observation with one-eighth the total number of observations (number of families observed per day times number of days).

Between 1961 and 1963 the Laboratory of Economic-Statistical Research of the Armenian Institute of Economics and the Laboratory of Mathematical and Statistical Methods of NII-truda in Moscow conducted a number of small-scale experimental budget surveys using the method of momentary observations in Yerevan and Leninakan (Armenia) and in the Bauman raion of Moscow (Vagradyan 1962, Levin 1974 p.198, Bauman 1970). The sample size was just 100 in Armenia and 500 in Moscow. The programme, however, was very detailed, covering family composition, living conditions, possession of consumer durables, holidays and consumption of free social services as well as income and expenditure (Balansy... 1969). Virtually nothing has been published on the results of these surveys. No doubt they were not substantial enough to yield clear conclusions. The economist Shvyrkov made the cautious judgement in 1965 that further research into the method was needed.

A much larger-scale and very successful experimental survey was conducted in Riga, Latvia, between October 1967 and September 1968. About 8700 families (according to Venetskii and Bauman (1975), though Vitols (1972) gives 8800) were interviewed, or some three per cent of the population of Riga, at the rate of 24 new families per day, by students from the Faculty of Economics of the Latvian State University named after P.Stuchka. Each family was interviewed on two successive days about their expenditures on the day of the interview, though some of the 160 questions on the questionnaire related to incomes and expenditures over the past month (Venetskii 1968). The sample was designed on the territorial principle. Extensive data analyses were carried out by punched-card equipment and by computer, and fairly informative reports of the results have been published (Berzkaln 1968b, Balansy... 1969, Krastin' and Berzkaln 1972, Venetskii and

Bauman 1975).

It will help us gain some idea of the internal politics of budget survey reform if we review the roles played by different organisations in the conduct of this survey. It was organised principally by the Scientific Research Institute of TsSU USSR, under the control of the Institute Directorate, with both researchers in the Sector of Sampling Methods at the Moscow headquarters (Marina Bauman, S.P.Ananyeva, L.A.Nikolayeva) and specialists in the Latvian Division of the Institute (O.Ya.Berzkaln, B.Ya.Mezhgailis) working on the project. These people, possibly with the exception of Ananyeva, are clearly educated in sampling theory. Two academic statisticians also directly participated in the work - I.Venetskii of the Moscow Economic-Statistical Institute and Ya.P.Vitols of the Latvian State University, students from which were, as we have noted, used as interviewers. The only clear contribution of TsSU itself, as distinct from the TsSU research institute, was that the facilities of the Computer Centre of the Latvian TsSU were made available to process the survey data. Venetskii (1968) admittedly lists as a co-organiser of the survey the Budget Statistics Department of TsSU USSR and as a participant in the research, second after Venetskii himself, its head Matyukha. However, no mention of this is made in any of the other articles on the survey, not even in a later article co-authored by Venetskii (Venetskii and Bauman 1975), nor does Matyukha make any reference to his alleged participation in any of his many works. We suspect that the participation of the Budget Statistics Department is a fiction invented for political reasons.

It is true that Matyukha takes a much less hostile stance on the method of momentary observations than he takes on the question of the territorial reconstruction of the budget survey. While he ignores or argues against the latter proposal, he praises work on the method of momentary observations as a "promising direction" for future research (Matyukha 1962), though

there is no sign that he contemplates adopting the method in practice. However, an unwillingness to change the existing budget survey can be combined with recognition of the method of momentary observations by regarding the latter as a possible supplement to the survey in its present form. Unlike the researchers associated with the method (Karapetyan 1962, 1980 p.253; Berzkaln 1968b, Venetskii and Bauman 1975), Matyukha does not see a system based on momentary observations as a potential full or partial replacement for the existing survey. Such an evasion is hardly possible on the issue of the territorial principle.

Between 1968 and 1971 another budget survey using the method of momentary observations was conducted in three towns of Kostroma oblast' (RSFSR) with a sample of about 6100 families designed on the territorial principle (Arkhipov 1974, Levin 1974 pp.195-201). The survey was organised by the Institute of Sociological Research which had been set up under the Department of Propaganda and Agitation of the Kostroma oblast' Party Committee in 1965, with the participation of the Department of Living Standards of the Academy of Sciences Institute of the International Workers' Movement and of the Laboratory of Models of Living Standards of TsEMI. The questionnaire covered not only income and expenditure, but also possession of household property, living conditions, availability of everyday services, unsatisfied consumer demand, attitudes to work and the use of non-working time. Unlike the survey in Riga, the Kostroma survey was not conducted for the purpose of methodological experimentation but in order to assess the socio-economic requirements of the population. Thus Levin uses data from the survey to analyse the problem of consumer goods shortages.

Although the Riga and Kostroma surveys appear to have proven successful, the 1970s have not seen further developments in this field. To some extent this may be due to the departure of key figures from the field. Karapetyan died in 1977. The interests of Rimashevskaya and her colleagues at TsEMI

have started to move away from family budget studies to such areas as health economics, though Rimashevskaya may now be supervising a new postgraduate student doing research on the method of momentary observations (Shenfield 1982b). Venetskii too is now dead. If these events have made some difference to the immediate prospects for change, this is of course only one more pointer to the extreme inertia with which the Soviet statistical system resists innovation.

5 The Riga survey - the sample design

The sample design of the Riga survey casts no special light on the method of momentary observations but is of interest as an indicator of the level of competence in sampling theory attained by the TsSU research institute at the time. The main accounts are those of Venetskii (1968) and Vitols (1972).

A two-stage proportional stratified sample of residential buildings (zhilye doma) was first selected by what appears to have been a sound procedure. Riga is divided into five raiony, each of which had been divided for the purpose of the 1964 youth census into neighbourhoods (schetnye kvartaly). A list of all the buildings in each neighbourhood was available from this census. One in ten of these lists was selected systematically, with allocation among the raiony proportional to the number of families in each raion. Buildings were selected systematically within each neighbourhood list, separately in three strata - buildings belonging to the City Soviet, buildings belonging to State establishments and Departments, and houses belonging to private persons. (This is according to Venetskii; Vitols reports only two strata, with the first two categories of building in a single stratum.) Data from the residential administrations were used to achieve proportional allocation of buildings among strata. The proportion of buildings selected is not made clear, but

must have been very large in view of the fact that about three per cent of the population were covered in a survey conducted in ten per cent of neighbourhoods.

In selected buildings containing not more than twenty families, all resident families were included in the survey. In buildings containing more than twenty families, twenty families were selected on different floors. Vitols tells us that the number twenty was chosen because new blocks of flats in Riga typically have twenty flats per entrance (pod'yezd), so presumably only one entrance in any block was covered. It is not clear how the choices of entrances and flats to be covered in large blocks were made, but this may have been left to the discretion of interviewers[2].

After three months of the year-long survey, the representativeness of the sample selected up to that point for interview was checked by comparing sample indicators with corresponding population indicators, and the results were apparently satisfactory. However, as we are not told which indicators were checked and as the range of population indicators available must have been limited, this does not exclude the possibility of bias on other indicators. The sampling procedure is susceptible to bias in a number of respects. If, for example, interviewers were allowed themselves to select flats in large blocks, they may have over-represented lower floors and more accessible entrances. As there was no provision for call-back when residents were out at the time the interviewer called, such families were not represented, though they were probably a group unrepresentative of the population as a whole (for example, able to afford eating out and travelling more frequently than average, more frequent visitors to cinemas, concerts etc.). The most important bias may be that the procedure gives a much lower probability of selection to families living in very large buildings than it gives to families living in small ones, including privately owned houses. Families in large buildings may well differ from

families in small ones on important variables such as occupation. There is also the problem of weighing up the practical convenience of concentrating the sample in clusters within blocks of flats against the loss in sampling efficiency entailed by the high intra-cluster correlation within such clusters; this is why Volkov comes out against clustering at this level in his dissertation (1971) on population sampling.

Vitols gives as the main motivation for the sample design the fact that it provides for an even territorial distribution of the sample. This is not of course necessarily desirable: it matters more that the sample be distributed evenly over the population, which may have a very uneven territorial distribution if tall blocks of flats predominate in some areas and low private houses predominate in others. At any rate, an even distribution of the sample is consistent with many possible sample designs and so cannot justify the choice of any particular design.

Thus the sample used in the Riga survey was not very well designed. It is difficult to judge to what extent this reflects on the understanding of sampling theory by the researchers, as they may have been constrained by practical considerations and may also have regarded the sample design as a matter of secondary importance in a survey with the main aim of studying the method of momentary observations.

6 The Riga survey - data analysis and findings

A variety of analyses are known to have been carried out on the Riga survey data, such as the fitting of empirical distributions of family income and of per-capita income, the calculation of consumption norms by age and sex, the calculation of income elasticities of consumption and the construction of mathematical models of income formation (Venetskii 1968). Most information is available on those analyses which were purely methodological in character, aimed at assessing the performance of the

method of momentary observations - comparisons with data from the continuous budget survey, and the analysis of sampling error.

Average indicators from the experimental survey were compared with corresponding average indicators from the TsSU continuous survey of 172 workers' family budgets in Latvia (Berzkaln 1968b). The account of the discrepancies found is not very clear, and it is strange that Berzkaln takes the continuous survey with all its biases as a standard for assessing the reliability of the experimental survey. Nevertheless, the number of purchases recorded by momentary observation of such irregularly purchased goods as radio and television sets (16), furniture (25) and cars and motorcycles (0) is so low that one suspects distortions unconnected with sampling error to be at work (figures given by Krastin' and Berzkaln (1972)). It is not surprising that expenditure on cinema, theatre and other cultural needs should be grossly underestimated, as few of those consuming such services on a given day would be at home to be interviewed about it!

The first calculations on sampling errors and necessary sample sizes are reported by Berzkaln (1968b) and Krastin' and Berzkaln (1972) of the Latvian Division of the TsSU research institute. Berzkaln first used the empirical method of examining successively larger cumulative subsamples to determine at what sample size different indicators became "relatively stable" (Table B11.2(a)). He stopped at a sample size of 800, although many budget indicators, as well as correlation and regression coefficients, were still unstable at this point. We are not told the criterion of "relative stability". The value of the exercise was also limited by the fact that it was carried out only once.

More theoretical calculations were also made of the sample sizes necessary to keep sampling error within 5 per cent and 10 per cent limits with 95 per cent probability in the continuous survey (using data from the Latvian workers' sample) and by momentary observation with two-day periods

TABLE B11.2

NECESSARY SAMPLE SIZES IN BUDGET SURVEYS USING THE METHOD OF MOMENTARY OBSERVATIONS

(a) Sample sizes at which different indicators become relatively stable

Indicator	Necessary sample size
Average monthly income	400
Expenditure on food	400
Expenditure on bread	600
Expenditure on meat	500
Expenditure on milk	400
Expenditure on sugar	300
Expenditure on fish	600
Expenditure on potatoes	still unstable at 800

Method: Random subsamples were accumulated from the data of the 1967-8 Riga budget survey using the method of momentary observations, starting with a subsample of 200 and proceeding by steps of 100 to a subsample of 800.

Source: Berzkaln (1968b)

of observation (Table B11.2(b)). Berzkaln concluded that samples of 2-3000 were adequate for fairly regular expenditures like those on milk and meat. Krastin' and Berzkaln showed how necessary sample sizes could be reduced by increasing the period of observation to a week or a month (Table B11.3). They devised a "rational plan of enquiry" with the period of observation for different indicators differentiated according to frequency of purchases (Table B11.4). Expenditures on some foods and petty non-food goods would be recorded for one day only, expenditures on other foods for one week, on clothing and alcoholic drink for one month, on consumer durables for one year and on vehicles for five years.

None of these early calculations were reliable, inasmuch as they used the formulas appropriate to simple random samples, ignoring the complex structure of the real sample. It is not clear whether the same is true of the calculations of sampling errors reproduced in Balansy... (1969) (Table B11.5). It appears that new calculations were later carried out by computer on a range of the more important indicators. The account given of them by Venetskii and Bauman (1975) indicates that they were well founded theoretically, probably because more competent mathematical statisticians like these authors had now become involved. Unfortunately new data on sampling errors and necessary sample sizes to replace the unreliable old data do not seem to have been published.

The new analyses were carried out in five stages. First, tests were carried out to determine whether the families surveyed in different time periods (days, weeks etc.) had been drawn from the same population. These tests showed an absence of time biases in the sample: there was very little variation in the socio-economic structure of subsamples interviewed in different months and quarters of the year[3]. Second, smoothed time-trends of indicators over a week, a month and a season were fitted, so that the fluctuation of transactions over time could be taken into account as well

TABLE B11.2

NECESSARY SAMPLE SIZES IN BUDGET SURVEYS USING THE METHOD OF MOMENTARY OBSERVATIONS

(b) Sample sizes necessary to keep sampling error within given limits in continuous budget surveys and in budget surveys using the method of momentary observations

Sampling error limit	Indicator	Sample size necessary in a continuous survey	Sample size necessary in a survey using the method of momentary observations	Ratio of necessary sample sizes
(col. 1)	(col. 2)	(col. 3)	(col. 4)	(col. 5) = $\frac{\text{(col. 4)}}{\text{(col. 3)}}$
± 5 per cent (p = 0·9545)	Expenditure on milk and milk products	440	2685	6·1
	Expenditure on meat and meat products	600	5080	8·5
± 10 per cent (p = 0·9545)	Expenditure on milk and milk products	111	674	6·1
	Expenditure on meat and meat products	140	1116	8·0

Note: Calculations based on data of the continuous FBS and of the budget survey using the method of momentary observations, relating to Latvia in 1967-8.

Source: Berzkaln (1968b)

TABLE B11.3

REDUCTION OF NECESSARY SAMPLE SIZE WITH INCREASING REPORT PERIOD IN THE RIGA BUDGET SURVEY USING THE METHOD OF MOMENTARY OBSERVATIONS

Product or product group	Number of purchases per day per 1000 families	Necessary sample sizes with a report period of		
		one day	one week	one month
Milk	667	1366		
Food products	x	1690		
Sugar	141	1739		
Meat	533	2018		
Fish	100	2051		
Bread	100	3053		
Eggs	61	17574	2511	
Potatoes	98	40385	5769	1346
Non-food products	948	53786	7684	1793

Note: The sample sizes were calculated so that the sampling error of the average expenditure on the given product did not exceed 5 per cent with probability 95 per cent. The formula used in the calculations was that for a simple random sample, ignoring the complex cluster structure of the sample actually used. The necessary sample sizes given in the Table are therefore under-estimates.

The source notes that the necessary sample size for some types of non-food products (presumably for such infrequently purchased items as cars) for a report period of one day exceeds 100 000.

Source: Krastin' and Berzkaln (1972 Table 2 p 38)

TABLE B11.4

A "RATIONAL PLAN OF ENQUIRY" FOR THE RIGA MOMENTARY BUDGET SURVEY OF 4000 FAMILIES BASED ON THE RELATION BETWEEN NUMBER OF REPORTED PURCHASES AND REPORT PERIOD

Product	Approximate number of recorded purchases by 4000 families for a report period of:				
	1 day	1 week	1 month	1 year	5 years
Purchase of food products					
Wheat bread	2136 (1)				
Rye and other bread	1968				
Potatoes	394	2758			
Vegetables and melons	1083				
Fruit and berries	605	4235			
Meat and salt	2133				
Fish and fish products	400	2800			
Milk	2668				
Smetana and cream	822	5754			
Tvorog	488	3416			
Animal fat	1094				
Eggs	243	1701			
Sugar	565	3955			
Purchase of alcoholic drink					
Vodka (2)	89	623	2670		
Wine	57	399	1710		
Beer	78	546	2340		
Purchase of non-food products					
Clothing, cloth, footwear	358	2506	10740		
Fabrics	27	189	810		
Footwear	48	336	1440		
Ready clothing, linen and headwear	108	756	3240		
Knitted goods, socks and stockings	107	749	3210		
Cultural products	980	6860	29400		
Radios and TV's	4	28	120	1460	
Furniture and household utensils	52	364	1560		
Furniture	6	42	180	2190	
Cars, motorcycles and bicycles	0·9	6	27	329	1645

TABLE B11.4 (cont'd)

Source: Krastin' and Berzkaln (1972 Table 4 pp 41-2)

Notes: (1) Figures corresponding to the optimal choice of report period are underlined.

(2) These figures are of course hardly credible, even for Latvia.

TABLE B11.5

COMPARISON OF SAMPLING ERRORS OF ESTIMATES OF MONTHLY EXPENDITURES ON FOOD PRODUCTS DERIVED FROM CONTINUOUS AND FROM MOMENTARY BUDGET OBSERVATIONS

Food product	Mean monthly expenditure	Mean sampling errors absolute		Mean sampling errors relative		Ratio of mean sampling errors
		CO	MO	CO	MO	CO : MO
	(1)	(2)	(3)	(4)	(5)	(6)
				$\frac{(2)}{(1)}$	$\frac{(3)}{(1)}$	$\frac{(2)}{(3)} = \frac{(4)}{(5)}$
	old rubles	old rubles		%	%	
Bread	81·5	10·1	2·9	12	4	3·5
Cereal and macaroni products	23·4	3·4	2·6	14	11	1·3
Vegetables	46·2	9·2	4·2	18	9	2·2
Fruit	41·4	8·8	4·7	20	12	1·9
Milk	42·0	6·6	2·6	15	6	2·5
Butter	61·2	9·1	5·2	14	9	1·8
Meat	122·8	16·3	7·3	13	6	2·2
Fish	22·9	4·4	3·0	18	13	1·5
Sugar	55·6	6·6	5·1	12	9	1·3

CO continuous observations

MO momentary observations

Source: Balansy... (1969 p 66); column (6) calculated by us

Method: The calculations are based on the expenditure records over 30 days of 200 families of workers in Latvia participating in the continuous family budget survey. These records probably relate to a month in 1967. The CO estimates were derived from data for individual families, the MO estimates from "diagonal data" (see Table B11.1 (d)) taken across individual families but including the same number of daily observations.

as their average frequencies. Third, the correlations among daily observations within two-day and three-day periods of observation were estimated, and analyses of variance into "inter-moment" and "intra-moment" components made. This made it possible to correct estimates of sampling errors for the interdependence of consecutive observations of the same family. Fourth, intra-stratum correlations were estimated, another necessary element in the final calculations to estimate sampling errors[4].

As a result of their analyses Venetskii and Bauman classified budget indicators into four groups in accordance with their patterns of fluctuation over time (Table B11.6) and their general suitability for study by the method of momentary observations (Table B11.7).

Group 1 covered those daily expenditures on food which showed insignificant fluctuation over time, so that sampling error depended only on the total variance and not on the inter-moment and intra-moment variances. Relative sampling errors estimated from data "for the majority of surveyed families" were not more than 5-6 per cent. Even one-day observations on the small daily samples (24 families) would have given estimates of average expenditures with sufficient precision.

The various expenditure and income indicators in Group 2 showed less regularity, with variation between days differing from variation between families interviewed on the same day. Though variation in time was not very striking, it was enough to prevent the Riga sample giving sufficient precision. Either the daily sample size or the period of observation (or both) would have to be increased, though sampling efficiency would be reduced by an unnecessarily long period of observation.

Group 3 consisted of food purchases showing a strong weekly cycle, with most expenditure concentrated on one, two or three days a week, depending on the good and the population point. Thus in Riga most fruit and potatoes are available for sale at the weekend, and most vegetables on

TABLE B11.6

DAILY VARIATION OVER THE WEEK IN SIZE OF FOOD PURCHASES

Food products	CV	Distribution of purchases over week							
		Mon	Tue	Wed	Thu	Fri	Sat	Sun	Total
	%	%	%	%	%	%	%	%	%
Group 1									
Bread products	3	15	15	14	14	14	15	14	100
Milk	5	15	15	14	14	14	14	14	100
Smetana and cream	8	15	14	14	13	13	16	15	100
Butter	8	15	14	14	13	15	16	13	100
Group 2									
Meat and fat	10	13	13	16	13	15	16	15	100
Fish	18	15	10	17	13	12	18	14	100
Sugar and confectionery products	19	13	13	15	10	13	20	16	100
Eggs	25	10	15	13	18	21	12	12	100
Group 3									
Fruit and berries	37	10	9	11	11	16	24	19	100
Potatoes	46	14	12	11	10	14	8	32	100
Vegetables and melons	47	9	30	12	10	11	13	15	100

CV coefficient of variation

Source: Venetskii and Bauman (1975 p 26), based on data from the 1967-8 Riga survey on the momentary method

TABLE B11.7

CLASSIFICATION OF BUDGET INDICATORS ACCORDING TO THEIR SUITABILITY FOR OBSERVATION BY THE METHOD OF MOMENTARY OBSERVATIONS

Group 1 - The sample size used in the Riga survey gave satisfactory results even with a one-day period of observation. Suitable for momentary observation.

expenditures on bread products, milk, milk products

total daily expenditure on food

Group 2 - The sample size used in the Riga survey gave inadequate results with a one-day or two-day period of observation. Results would be satisfactory if the sample size were larger or the period of observation increased to three days. Suitable for momentary observation.

expenditures on meat, fish, eggs, sugar and confectionery items, fats and oils

expenditures on specific categories of these products (eg mutton, veal, pork, animal fat, sugar)

daily expenditures on services and other non-goods

expenditures on specific categories of these services (eg cinema and other spectator entertainments)

monthly incomes (pensions, grants and other State payments, alimony, help from relatives)*

Group 3 - Because of the cyclical pattern of purchases (see Table B11.6) the period of observation must be one week. Suitable for momentary observation.

expenditures on potatoes, vegetables and melons, fruit and berries

Group 4 - Because of irregularity of transactions a very long period of observation is necessary. Not suitable for momentary observation.

daily and monthly expenditures on non-food goods, especially on consumer durables

various types of daily and monthly money receipts

some types of non-goods expenditures

* ie reliable information can be obtained from interviews on one or two days only.

Source: Venetskii and Bauman (1975)

Tuesday. Where such periodicity exists, sampling error is minimised by observing families throughout one whole cycle.

Group 4 consisted of infrequent and irregular transactions subject to very large sampling errors when momentary observation is used and for which the Riga survey did not give reliable data. It included consumer durables, vehicles and various types of money receipt[5]. Much longer periods of observation are here necessary.

Venetskii and Bauman conclude that Groups 1-3, comprising about one-half of all budget items including expenditures on basic food products and basic money incomes, are characterised by sufficient frequency and regularity to make it expedient to transfer them from the continuous budget survey to a new territorially based survey using the method of momentary observations. This will save time, resources and labour. One-day periods of observation are recommended for Groups 1 and 2, one-week periods for Group 3 with daily partial renewal of the sample. The periods of observation required for Group 4 are so long that these items may as well remain covered by the continuous budget survey. Bauman (1970) states that momentary observation is also unsuitable for the study of living conditions and demographic indicators in view of their long-term stability.

Berzkaln (1968b) makes the further suggestion that more detailed information on population subgroups of special interest - families of workers and employees with private plots, families with very low incomes, families with very high incomes etc. - be separately collected from subsamples with a period of observation of one year.

7 Conclusions

The use made of the branch principle and the excessive periods of participation are properly regarded by Soviet critics of the FBS as the two main defects of its organisation. The proposals for the reconstruction

of the survey on the territorial principle, examined in the last chapter, represent basically a competent attempt to adapt the methodology of Western, and in particular American, social surveys to Soviet conditions. The proposals for reducing the period of participation, examined in this chapter, represent more than imitation of the West. The application of the method of momentary observations, taken from the field of industrial work study, to budget surveys is an imaginative contribution to the theory and practice of budget surveys which would surely attract considerable professional interest in the West were it more widely known there. Its introduction into practice would involve a break in Soviet tradition more radical than would be involved in, say, the adoption of the British practice of a standard one-year period of participation. This, unlike the method of momentary observations, would at least retain the principle of collecting balanced individual budgets of income and expenditure. Although some of the work on the Riga survey was of uneven quality, this and other experimental surveys have confirmed the practical feasibility of the method of momentary observations and laid the basis for further advance - which, however, has not as yet come.

The implications for the problem of methodological innovation in Soviet statistics seem to us twofold. It is clearly not the case that the Soviet system always prevents researchers from pursuing original research, as distinct from mere imitation of Western work. They may sometimes be able to obtain the resources to develop their ideas by means of quite large-scale experimental exercises, to publicise their views and to obtain academic and even some degree of official recognition. But equally clearly, new ideas meet a formidable barrier to their progress at the point of introduction into general practice. Even the formation of an academic consensus in their favour and their formal advocacy by the research institutes attached to TsSU itself do not suffice to motivate TsSU officials

to change long-established practices. Indeed, the method of momentary observations has not been very widely introduced even in its original field of application, industrial work-study, where it was tried out in the USSR as early as 1924.

Possible explanations for this methodological conservatism will be considered in Part E.

Notes to Chapter B11

1 The State Planning Commission was at this time divided into two agencies - Gosplan for short-term planning and Gosekonomsovet for long-term planning. NIEI now again comes under a united Gosplan.

2 Hostels and Army accommodation were excluded from the Riga survey.

3 The coefficients of variation over time of such indicators as the proportions of families belonging to different social groups, the average size, wages and per-capita income of families in different social groups, and the distribution of families by numbers of members, did not exceed 8-9 per cent.

4 The indicators studied were: money income per month, wages per month, daily quantities of food products purchased, expenditure on non-food products, total expenditure on food, non-goods and other money expenditures. The statistics calculated for each indicator for each month included: averages, variances between days (inter-moment), variances between families (intra-moment), ratio of inter-moment to intra-moment variances, chi-squared tests, coefficient of association, coefficient of intra-stratum correlation, absolute and relative sampling errors.

5 Unreliable results were obtained on irregular money receipts not only because of the unevenness of their distribution over time but also because, so it is thought, respondents gave inaccurate and incomplete information (Berzkaln 1968b, Bauman 1970). Berzkaln suggests that an effort be made to collect more precise data on "other incomes" from a 1-2 per cent sample.

CHAPTER B12

THE SAMPLE DESIGN OF THE FBS AND THE LEGACY OF THE PAST

1 Introduction

In Chapter A4 we discussed how the arrested development of sampling in the USSR has affected the contemporary practice of TsSU, both by preserving the use of non-probabilistic forms of sampling and by contributing to deficiencies in the conduct of probabilistic sampling. In Part B we have noted at various points the survival in the sample design of the FBS of features characteristic of the formative period of sampling, such as reliance on comparisons of sample and population means in representativeness checks (Chapter B6) and the predominance of the branch over the territorial principle (Chapter B10). We have also seen how outdated conceptions influence attitudes concerning the optimal size of the FBS sample (Chapter B2). In this chapter we make an overall assessment of the extent to which the sample design of the FBS can be explained as a legacy of the past.

Those Soviet writers who recognise the existence of outdated elements in the sample design lay great stress on the influence of the tradition of monographic budget surveys dating back to nineteenth-century Russia. We describe the monographic origins of budget surveys in Section 2. In Section 3 we consider the three main Soviet points of view on the "nature" of the FBS: that of the TsSU officials who claim that it is organised in accordance with modern sampling theory; that of the writers who criticise its "monographic" aspects; and that of adherents of the monographic school of thought, who criticise it for being insufficiently monographic. In Section 4 we assess the likelihood that various features of the sample design, such as the use of the branch principle and long periods of

observation, are of monographic origin. We conclude in Section 5 with a discussion of the nature and importance of historical determinants of the sample design.

2 The monographic origin of budget surveys

The study of household budgets was introduced in Western Europe by the social investigator LePlay. On a trip through Germany in 1829, he stayed with a miner's family and recorded detailed observations on their way of life (Lazarsfeld 1961). Over several decades he collected hundreds of such "monographs". The detailed nature of these accounts, discursive as well as quantitative in format, is shown by the fact that the 58 monographs published in the 1877 edition of "The European Workers" occupied six volumes. Each monograph contained a description of the locality and occupation, institutions of apprenticeship, contracts etc, the history of the family, their style of life and sources of subsistence, as well as data and commentary on income and expenditure.

This method of studying a few families in great detail, which came to be known as the "intensive" or "monographic" method of investigation, was first used in Russia by D.P.Zhuravskii, who published in 1846 data on the budgets of two families with different income levels. His purpose was to compare the patterns of expenditure typical of different social classes.

In the last quarter of the nineteenth century, the statistical bureaux of many zemstva (bodies of local self-government created in 1861) conducted monographic studies of peasant households. In the first of these studies, conducted in Ryazan' guberniya in 1877-8, twelve budgets were compiled in all, with two budgets representing each of six "types" of peasant household, ranging from "rich" to "indigent" (Matyukha, Postnikov and Samoilov 1958). Larger-scale surveys of peasant households followed in the 1880s and 1890s, to be supplemented after the turn of the century by

some surveys of households of industrial workers (see Table B10.4).

The most common, though not the only, principle used in selecting households for the early monographic surveys was that of "typological" sampling. It was assumed that the population consisted of relatively few distinct "types" of household, identified in advance in accordance with theoretical considerations, and that differences among households of the same "type" were not significant. Therefore it was sufficient to select a very few, perhaps just one or two, "typical" households as representatives of each type. Types were defined in terms of social class, economic conditions, locality, branch of industry and so on. For example, it was proposed at the Russian Statistical Congress in 1894 that the peasant correspondent of the statistical bureau in each area be requested to provide budget data on a household he considered to be typical of his area (Wheatcroft 1980, Vol.1 p.34). The average size and characteristics of the selected households were then to be compared with census data.

Thus, putting aside the difficult problem of how to make sure of obtaining a "typical" sample, it seems that the ideal of typicality being aimed at was a close correspondence between <u>each</u> member of the sample and the average of the type population with respect to important known variables. It was wrongly assumed that if the units selected were typical with respect to known variables, they could also be taken as typical with respect to the unknown variables being investigated (O'Muircheartaigh and Wong 1981). The assumption of homogeneous types led to oversight of the desirability of representing the variability of the population in the sample. Matyukha (1967 p.162) tells us that the statisticians of the Penza <u>zemstvo</u>, in selecting "typical" peasant households from "typical" villages, set themselves the task, according to the <u>Statisticheskii vestnik</u> of 1914, "of guaranteeing that nothing atypical or exceptional be included in the survey". Matyukha's approving comment that "this had great significance in providing

for the representativeness of the sample" shows the continuing influence over concepts of representativeness of the old monographic concept of typicality.

3 Soviet points of view on the "nature" of the FBS

We saw in Chapter A4 that in the Soviet classification of "forms of statistical observation" a distinction is made between "incomplete" (nyesploshnye) surveys and "sample" (vyborochnye) surveys proper. Incomplete surveys are all those based on incomplete enumeration of the population, irrespective of the method of selection used, while only surveys designed on the basis of probabilistic sampling theory are counted as sample surveys proper. The distinction is made necessary by the continued use of various forms of non-probability sampling in Soviet statistics.

As the FBS sample is designed only partially in accordance with the dictates of probabilistic sampling theory, there is room for controversy regarding the definition of the FBS as a "form of statistical observation". There is among Soviet writers a dispute about the "nature" of the FBS which, in spite of its occasionally scholastic appearance, concerns both the adequacy of the FBS sample and the historical origins of the sample design. It is useful to distinguish three basic points of view taken in the dispute:

(a) The point of view of TsSU officials who recognise that sampling proper is "the scientifically most perfect form of incomplete observation" and who also designate the FBS as a sample survey proper, thereby implying (if not explicitly stating) that the sample design satisfies the requirements of sampling theory (Matyukha 1962, Ananyeva 1964 p.31)[1].

(b) The point of view of critics who deny that the FBS can be considered a sample survey "in the full sense", on the grounds that it fails to satisfy the "basic requirements of the sampling method"

(Vagradyan 1962). Such critics often designate the FBS in a way that draws attention to the alleged retention in the sample design of monographic elements - for example, Karapetyan and Rimashevskaya (1959) describe the FBS as "a monographic sample survey". They argue for the elimination of these elements.

(c) The point of view of certain statisticians who recognise the existence of monographic elements in the FBS sample design, but who argue not for their elimination but for their accentuation and for abandoning the elements of probability sampling. "(The collection of) family budgets must not be regarded as a sample survey; this is not a sample but detailed monography" (Maslov 1967 p.155).

Let us take these points of view in turn.

(a) The apologetic point of view

We have seen that there is a tendency in official accounts of the FBS to deny, conceal or minimise the divergence of the sample design from the requirements which modern sampling theory makes of a sample survey. Thus, the false claim is made that coverage by territory and branch is proportional to population (Chapter B1); the exclusion of unskilled workers from the survey is concealed, as is that of collective farmer households which share a cow with other households (Chapter B4); and the probably false claim is made that the response rate is almost 100 per cent (Chapter B5).

At the same time, one finds in the writings of TsSU officials a measure of ambivalence towards the sampling theory the requirements of which they recognise in theory and resist in practice. This often takes the form of unconvincing arguments to the effect that sampling theory has not been adapted to supposedly special Soviet conditions. There is "an increasing lag of sampling theory behind the requirements of practice... The classical

mathematical schemas for sample surveys are insufficient to solve the problems arising out of Soviet reality" (Starovskii 1962).

The book on the FBS by the former Head of the TsSU Budget Survey Department I.Ya.Matyukha (1967) is marked by an inconsistent attitude towards sampling theory, of which the author is probably not himself conscious. We have observed in Section 2 that in his account of the pre-revolutionary budget surveys Matyukha uncritically accepts that their monographic basis provided for representativeness. Elsewhere in the book, however, due obeisance is paid to modern sampling theory, as in the treatment of sampling error, which gives the false impression that sample sizes in the FBS are set on the basis of sampling-theoretic calculations.

(b) The modern point of view

We may take the views of A.Kh.Karapetyan and N.M.Rimashevskaya as representative of the views of those who deny that the FBS can be considered a sample survey proper while arguing that it should be transformed into one. It is a "monographic sample survey" (Karapetyan and Rimashevskaya 1959), in which "traditional forms of organising budget surveys with monographic origins" are retained (Karapetyan 1980).

A variety of arguments are advanced in support of this proposition. Karapetyan and Rimashevskaya (1959) declare that "the nature of any survey is determined above all by its programme. The FBS, in the wide coverage of its questions, is a monographic survey". Karapetyan (1980) considers the use of the branch principle and of long periods of participation, and even the practice of collecting complete budgets from individual families, as products of the monographic origins of the FBS - points which we shall assess in Section 4. The most substantial arguments, however, are those which stress the limited extent to which sampling theory is applied in the sample design. Thus, while in a sample survey the sample size is calculated to achieve a given precision of estimation, "in a monographic

survey (as in the FBS - SDS) the sample size may not be the subject of a calculation" (Karapetyan and Rimashevskaya 1959).

The most pertinent observation made by Karapetyan and Rimashevskaya (1959),and the one which motivates their term "monographic sample survey", is that the FBS sample design more or less satisfies the requirements of sampling theory within individual sub-populations, defined by social group, territory and branch, but not in the way different sub-populations are represented in the overall sample. Similarly, Vagradyan (1962) states that the FBS "has become to a certain extent a representative sample within individual branches". The sample is therefore representative only at the level of individual sub-populations. As we saw in Chapter B6, it is at this level that the standard representativeness checks are carried out. In the haphazard selection and coverage of individual sub-populations is reflected the influence on the FBS sample design of the early monographic budget surveys, which were restricted to particular territories and branches and were uncoordinated with one another (see Table B10.4). We shall return to this point also in Section 4.

(c) The point of view of the monographic school

We may take Professor P.P.Maslov and A.E.Zhidikhanova as representative of those writers who advocate that the FBS be transformed into a purely monographic survey (Maslov 1967 p.155, 1971 pp.35-6, Zhidikhanova 1982). In their view the aim of a budget survey should be not to achieve territorial representativeness but to explore differences among "types" of families. These differences can be described monographically on the basis of a small number of selected families because variation within each "type" is by assumption insignificant:

> There are narrow limits to variation in consumption. An adult needs about 3000 calories per day, and within a given occupation or social group fluctuations are small. Of course, the consumer has a great choice, but not so great by statistical criteria (Maslov 1967 p.155).

Systematic sampling is accordingly "open to criticism" and should be replaced by "selection by types", which would make possible a severalfold reduction in sample size. The reduced sample "would give basically the same results". (See Chapter B2.) However, those families which are observed should be studied in the greatest possible detail, and Maslov questions whether the existing FBS programme is sufficiently full. The influence of the ideas of nineteenth-century monography in all this is quite clear.

An assessment of the extent of the influence exercised by this openly monographic school of thought may help us judge the degree to which Soviet thinking on the FBS sample design is determined by the legacy of the past. For this purpose we shall briefly review the professional biographies of two of the most important representatives of the school, Professor P.P.Maslov and Ye.O.Kabo (Nekrolog 1978, Nekrolog 1970).

Professor Maslov (1902-1978) graduated from the Marx (later Plekhanov) Moscow Institute of National Economy and spent most of his working life lecturing in statistics at the Moscow Financial Insitute. In the 1930s, however, he organised demographic and agricultural censuses in Tuva and the Far North. He wrote widely on agricultural, financial and household budget statistics, and on the statistics of international trade, and for many years served on the scientific-methodological councils of TsSU USSR and TsSU RSFSR.

Ye.O.Kabo (1888-1968) was a Bolshevik activist in her youth but had no higher education. In the 1920s she headed the Department of Workers' Budgets of the Central Bureau of Labour Statistics, and then worked in research institutes in the 1930s and 1940s. Most of her publications deal

with budget studies, often in a historical perspective[2].

The fact that writers of the monographic school are in a position to have their work published at all argues that they are not completely without influence. Some statisticians who do not themselves actively propagate monographic ideas do not object to having their names associated with such ideas. For example, an article written jointly by Maslov and two other statisticians, K.P.Romanovskaya and V.M.Simchera, recommends that "the role of monographic methods of investigation" of FBS data be raised and complains that the proposals previously made by Maslov (1973) have been ignored (Maslov, Romanovskaya and Simchera 1978). On the other hand, Kabo's exposition of the theoretical rationale of monography, in which she takes issue with the criticism of monography made in 1910 by the famous Russian statistician A.A.Chuprov, is preceded by an editorial disclaimer drawing attention to the controversial nature of her ideas (Kabo 1972).

It seems likely on balance that the monographic school retains some influence, though a rather modest one. The very existence of the school must, however, make it more difficult for mathematical statisticians to make good their claim to theoretical authority in the eyes of "practical" statisticians apparently sceptical of the pretensions of both rival sets of theoretical statisticians.

4 The possibly monographic origin of various features of the FBS sample design

In this Section we consider whether various features of the FBS sample design can reasonably be attributed to the monographic origins of budget surveys in Russia:

(a) the incomplete and uneven coverage of different sub-populations defined by social group, territory and branch;

(b) organisation of the sample on the branch principle;

(c) prolonged periods of participation in the survey;

(d) the collection of complete budgets from individual families, as opposed to use of the method of momentary observations (Chapter B11); and

(e) certain practices which suggest an understanding of "representativeness" influenced by the monographic concept of "typicality".

Feature (a) is explained in terms of monographic origins by Karapetyan and Rimashevskaya (1959) and by Vagradyan (1962). That features (b), (c) and (d) are of monographic origin is argued by Karapetyan (1980).

(a) Incomplete and uneven coverage of sub-populations

As noted in Chapter B10, surveys of workers' budgets in pre-revolutionary Russia were restricted to "partial or local populations" of families associated with particular localities or with particular branches within localities (the Petersburg textile industry, the Baku oil industry etc.), and were "without pretension to any wide representativeness" (Vagradyan 1962). Although the budget surveys of the Soviet period were successively broader in scale and coverage, with originally local samples within particular branches being expanded in geographical scope, the coverage of sub-populations remained incomplete and uneven. This incomplete and uneven coverage of the population by the FBS sample can therefore be regarded as a legacy of the earlier budget surveys (Karapetyan and Rimashevskaya 1959).

However, the crucial factor in this development was presumably the local and uncoordinated nature of the earlier surveys. The fact that many of the pre-revolutionary surveys were designed on monographic principles does not seem pertinent to the argument.

(b) Organisation of the sample on the branch principle

Most surveys of workers' budgets in pre-revolutionary Russia were

organised on the branch principle (see Table B10.4). They covered local branch complexes at a time when whole families tended to be employed by a single industry from generation to generation and when each industry tended to be concentrated in a particular locality.

As the coverage of the FBS was expanded geographically, and as industrialisation did away with the restriction of branches to particular areas, the study of territorial-branch complexes was superseded by that of Union-wide branch breakdowns. For example, oil workers in Azerbaijan were now taken together with oil workers from new oil-fields on the Middle Volga (Tatar ASSR), in Western Siberia (Tyumen') and so on. This constituted an extremely heterogeneous group with respect to climatic and living conditions and consumption patterns, average data for which had little meaning.

Moreover, there is a growing tendency nowadays for members of the same family to be employed in different branches, or even to fall on different sides of the worker-employee division, so that very many families cannot be rationally allocated to any one branch or even social group (worker, employee). As a result the traditional methodology is outmoded (Karapetyan 1980 p.28).

The predominance of the branch principle seems, however, to be a specifically Russian tradition with no necessary connection with the monographic method as such. Early budget surveys in other countries did often cover particular occupations or industries, but often, as in France, they were organised on a territorial basis (International Labour Organisation 1925). The first surveys in the USA were of workers employed in various industries (iron, steel and allied industries in 1890, cotton, wool and glass industries in 1891), but these were soon followed (in 1901-2 and 1918-9) by surveys of workers' families in the main industrial centres selected without reference to industry (International Labour Organisation 1926).

That there is no necessary connection between monography and the use of the branch principle is confirmed by the fact that Maslov (1967 p.155) argues simultaneously in favour of reconstruction of the FBS on a monographic basis and in favour of abandoning the branch principle, recognising that "the structure of expenditures depends not on the branch in which the head of the family works but on real per-capita income".

(c) Prolonged periods of participation in the survey

Karapetyan (1980) describes the observation of "a restricted number of families over a long period" as a "monographic form". However, neither the early Russian budget studies (see Table B10.4) nor the early Western surveys usually extended over long periods. The practice of prolonged periods of participation seems to have arisen only in the Soviet period. It is conceivable that it is in part motivated by a desire to be in a position to follow changes in the budgets of particular families over long periods, in accordance with the monographic approach. Thus Aleshina (1965) describes the growth in living standards of the same 103 families over the eleven years 1951-1961, and Kabo (1967) regrets only that this comparison cannot be extended back to budget data from the 1920s. This work, however, is based on data from a small-scale budget survey conducted by the trade unions (VTsSPS); it could not be carried out on data from the TsSU survey because TsSU stores data from individual family budgets for only one year (Chapter C5).

Perhaps one could impute some "monographic" significance to the detailed propagandistic accounts of how life has changed for the better for individual, presumably typical, families in the budget survey which occasionally appear in the press or elsewhere (for example, Tatarskaya and Guryanov 1957).

(d) The collection of complete budgets from individual families

We have seen how use of the method of momentary observations would replace the collection of complete budgets for individual families, in which income balances expenditure, by the collection of different categories of budgetary data from independent samples with different periods of observation (Chapter B11). As an advocate of this method, Karapetyan (1980) argues that balances for population groups of similar families would be just as valuable as individual family budgets, and that:

> ... the compilation of individual family budgets has become a serious obstacle which must be overcome (above all, psychologically) in order to use properly statistical methods of mass observation.

Thus he regards the collection of complete individual budgets as another outdated practice carried over from monographic surveys. However, irrespective of the advantages and disadvantages of abandoning complete budgets, this is not essential for the application of "properly statistical methods": Western surveys of complete household budgets are genuine sample surveys.

(e) The concepts of "typicality" and "representativeness"

In Section 2 we made the distinction between the monographic concept of "typicality" and the modern concept of "representativeness". A "typical" sample is ideally one in which every unit is typical, or average, for the population or for a sub-population ("type"), while a "representative" sample ideally reflects the structure of the population in full, including its atypical parts. A number of features in the sample design of the FBS suggest that the concept of "representativeness" being used is to a certain extent distorted by the influence of the older concept of "typicality":

i. the use of mid-interval starting-points in systematic selection of enterprises, workers, collective farms, collective farmer households etc., leading to the "tail-cutting bias" which reduces sample variability (See Chapters A4 and B1);

ii. sole reliance on the comparison of sample and population means in checks of sample representativeness, with no attention paid in practice to comparing other aspects of the distributions of survey variables (See Chapters A4 and B6); and

iii. the method of directed selection of replacements for the purpose of improving the representativeness of the sample, conceived of in terms of means only, leading (as in i.) to the reduction of sample variability (See Chapter B8).

The case for attributing these practices, at least to some degree, to the influence of the monographic tradition is perhaps stronger than the case for so attributing such features of the sample design as incomplete and uneven coverage of sub-populations, use of the branch principle and prolonged periods of participation. However, even here the idea of monographic origins remains, in the absence of more thorough historical research, no more than a suggestive hypothesis[3].

The importance of those features of the sample design of the FBS which appear to reflect an outdated conception of representativeness should, moreover, not be exaggerated. The concept of "representativeness" underlying the sample design is on the whole a modern one. Thus, though the use of mid-interval starting-points reflects a notion of "typicality", systematic sampling itself is designed to achieve a sample representative of the entire population distribution on the ordering variable.

5 The legacy of the past: concluding discussion

The influence of outdated traditions on the sample design of the FBS is not very surprising in the light of the history of sampling in the USSR outlined in Chapter A3. The FBS was first established around 1930, just after the development of sampling in Soviet socio-economic statistics was frozen. At this time modern sampling theory had not yet been fully elaborated, let alone fully integrated into practice. The sample design of the FBS was accordingly only imperfectly informed by modern sampling theory. At no point since the 1930s has the FBS been fundamentally reconstructed in the way that would be required to bring its sample design fully into line with modern sampling theory.

The sample design of the FBS is in consequence rather eclectic in nature, incorporating both outdated and modern elements. Its transitional character is reflected in the variety of the criticisms to which it is subjected. Those who understand mathematical statistics criticise the FBS for its divergencies from the requirements of modern sampling theory, while those who remain committed to archaic sampling conceptions criticise it for conforming too closely to the requirements of modern sampling theory. If the sample design is far removed from that of a properly organised modern sample survey, it is also far removed from early forms of non-probability sampling such as monography.

In the West large-scale budget surveys were in general organised only after the Second World War, by which time modern sampling theory was fully developed and could be applied without difficulty. It was not necessary, as it would have been in the USSR, to undertake a basic reconstruction of already established budget surveys in order to modernise their sample design. The deficiencies of the present-day FBS are therefore in a sense the product of its relatively early establishment.

Although the nineteenth-century monographic tradition may well have

been one of the formative influences on the sample design of the FBS, it would surely be even more misleading to designate the FBS a "monographic" survey than it would be to designate it a sample survey proper. The importance of outdated elements in the sample design should not be exaggerated. Furthermore, by no means all features of traditional origin have any necessary connection with monography as such. Soviet writers such as Karapetyan tend to attribute to the influence of monography aspects of the FBS which have a traditional origin independent of monography (use of the branch principle) and even aspects which appear to lack traditional antecedents (prolonged periods of participation)[4]. The legacy of the past in the sample design of the FBS is a complex and varied one.

Notes to Chapter B12

1 We can also include under this heading the point of view of those who consider the FBS a sample survey as regards its sample design but not as regards its programme: "(The FBS is) monographic in the breadth of its programme, but a sample survey in the heterogeneity and size of its sample" (Shvyrkov 1965).

2 The biography of A.E.Zhidikhanova is not known. An article by her about the FBS, written from the monographic point of view, was published by the Central Mathematical-Economic Institute in 1982 (Zhidikhanova 1982).

3 One circumstance which casts some doubt on the hypothesis is the opposition of such representatives of the monographic school as Zhidikhanova to what they see as excessive reliance on means. They object especially to "mechanical" means which confound the various "types". They regard the identification of the "typical" as a subtle judgement requiring insight into the "essence" of the phenomena under study: "There is no guarantee that the genuinely typical should be reflected in the mean" (Zhidikhanova 1982).

4 The term "monographic" often seems to be used in a very loose sense to refer to any method of sample selection which is subjective in nature or does not accord with sampling theory.